Mads Corfixen, 2016 4715
mcorfi16@student.aau.dk

Data Visualisation Exam Paper
Visualisation of Pokémon!

Aalborg University
Data Science, 3rd Semester

1 Introduction

The interactive visualisations shown throughout the report can be found on this GitHub Pages site, where a copy of this report is also available.

1.1 Description of Used Data

The static dataset utilised is of the table type, containing 800 items pertaining to different Pokémon species (Pikachu, Bulbasaur, etc.), and 23 attributes which describe the different characteristics of each species. The dataset originates from [1]. The 6 attributes, which are used in this report, are shown below, along with an explanation of them and their type. Finally, an example of how the items and attributes of the dataset is structured can be seen in Table 1.

Type1 & Type2
The attributes Type1 and Type2 describe the typing of a Pokémon. Each species of Pokémon is defined to have a typing consisting of either one or two *elements*. These elements are Normal, Fire, Fighting, Water, Flying, Grass, Poison, Electric, Ground, Psychic, Rock, Ice, Bug, Dragon, Ghost, Dark, Steel, and Fairy [3]. If a Pokémon only has a single element as their typing, e.g. Pikachu, which is Electric, the entry of Type2 for that Pokémon is empty. Pokémon with only a single element are called mono-typed, and Pokémon with two elements are called dual-typed. Both of these attributes are nominal and qualitative as they are labels with no ordering.

Base Total
Not all Pokémon are equally strong. Each species is defined by their *statistic*, which assigns a numerical values describing their Hit Points, Attack, Defense, Special Attack, Special Defense, and Speed respectively [2]. The sum of these individual statistics is called the *base total* of that species of Pokémon. This attribute has values in the interval $I_{BT} = [180; 780]$ and the type of this attribute is interval and quantitative as it is possible to order and calculate differences between its values, but no zero value exists. The ordering direction is sequential.

Capture Rate
Not every Pokémon is equally easy to capture. This attribute describes the rate with which each Pokémon is caught, with lower values indicating more difficult catches. With values in the interval $I_{CR} = [3; 255]$, the type of this attribute is also interval and quantitative, and its ordering is sequential.

Is Legendary
Certain species of Pokémon are unusually stronger than all others. Such Pokémon are called legendary, and are usually used as the mascot for each Pokémon game. The type of this attribute is dichotomous qualitative, as it describes a boolean value; either a Pokémon is legendary, and is assigned the value of 1, or it is not, and is assigned the value of 0.

Base Egg Steps
Another aspect to each Pokémon is breeding. In the Pokémon games, every non-legendary Pokémon is capable of laying an egg, which is incubated by taking steps, i.e. walking, in the game. This attribute describes the exact number of steps needed for each Pokémon species' egg to hatch. With values in the interval $I_{BES} = [1,280; 30,720]$, the type of this attribute is also interval and quantitative, and its ordering is sequential.

Table 1: An example of how the items and attributes of the dataset are structured.

Name	Type1	Type2	Base Total	Capture Rate	Is Legendary	Base Egg Steps
Pikachu	Electric	None	320	190	0	2,560
Jigglypuff	Normal	Fairy	270	170	0	2,560
Mew	Psychic	None	600	45	1	30,720

Using these attributes, the tasks of interest, which this report and the visualisations contained herein seek to answer, can be defined.

1.2 Problem Statement

How can interactive data visualisations be used to solve the following tasks?

1. Which mono- and dual-typing are the most common?

2. Is there a correlation between the Base Total of a Pokémon species and its other attributes? If so, what is the correlation?

3. What makes a Pokémon legendary?

2 Development of the Program

In this section, ...

Written in Python...

Using these packages...

The structure is...

The complete program on GitHub...

Derived attribute Type...

3 Design Choices

In this section...

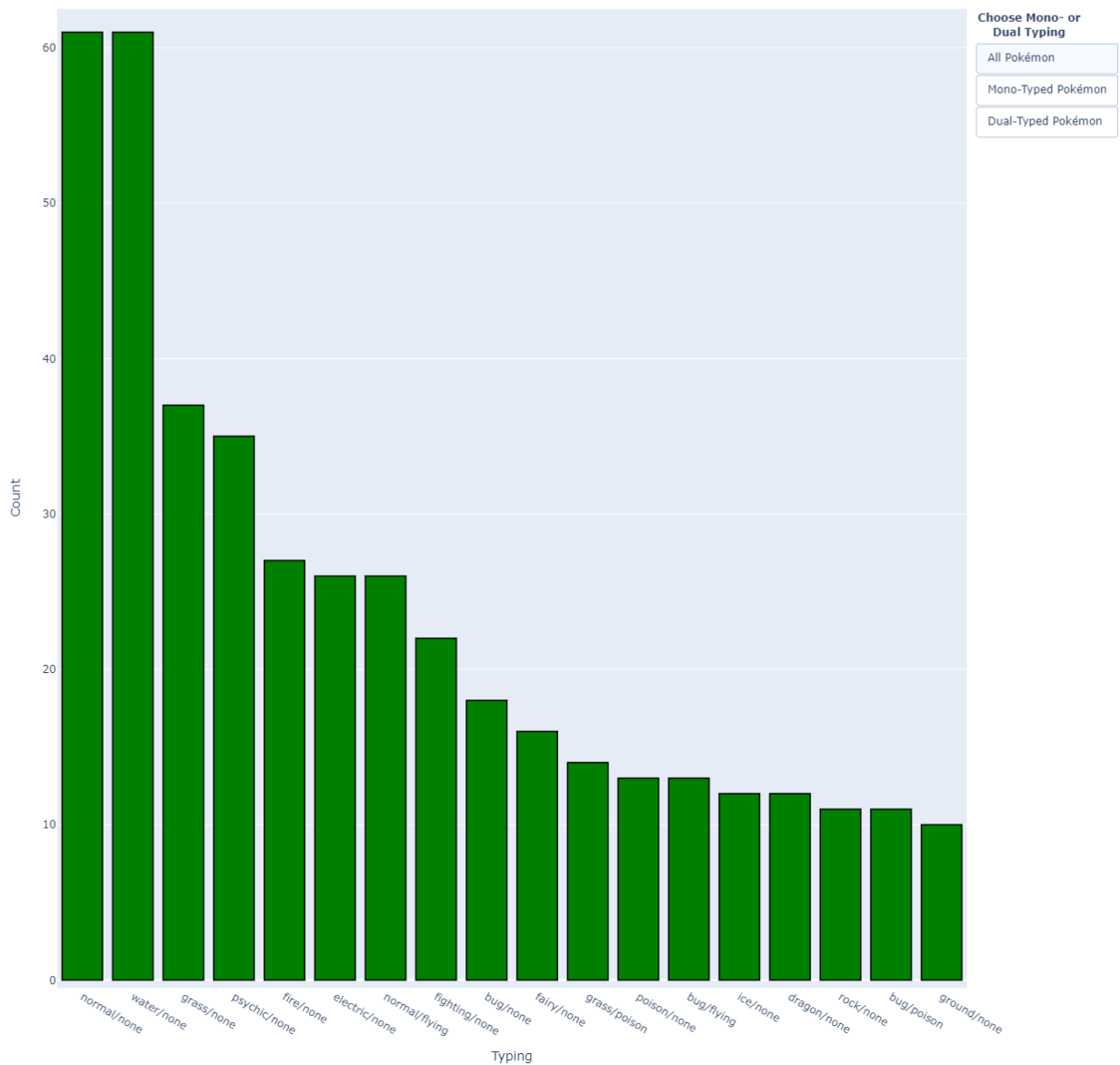Distribution of All Pokémon Types

Figure 1: Caption.

Figure 2: Caption.

Keywords:

- Symbols
- Encoding
- Annotations
- Items to Marks
- Attributes to Channels
- Idioms
- What, Why, and How

- Expressiveness Principle

- Effectiveness Principle

- Grouping

- Discriminability

- GESTALS Principles

- Labelling

- Responsiveness Categories

- Affordance

- Signifiers

- Feedforward

- Feedback

- Indicators

4 Usability Test

In this section...

4.1 Tasks

The respondent were asked to solve the following tasks using the visualisations.

1. Which mono-typing of Pokémon is most common?

2. Which dual-typing of Pokémon is most common?

3. Does the values of a Pokémon's Base Total and its Base Egg Steps seem correlated? If so, is it mostly linear, exponential, or logarithmic?

4. Does the values of a Pokémon's Base Total and its Capture Rate seem correlated? If so, is it mostly linear, exponential, or logarithmic?

5. What effect does it have on a Pokémon's Base Total, Base Egg Steps, and Capture Rate if it is categorised as Legendary?

4.2 Results

In Table 2, the feedback of the usability test is seen.

Table 2: An overview of the results of the usability test.

Task	Feedback
Task 1	1. 1
Task 2	2. 1
Task 3	3. 1

4.3 Changes Derived

5 Ethical Concerns

- Error margins (violin plots)

- Some Pokémon are better than others ⇒ Might homogenize Pokémon teams

- Remember 'Best Practices' from Slide 40 and 'Considerations' from Slide 55

- Zoomed-in as standard

6 Conclusion

References

[1] R. Banik. (2017). "The complete pokemon dataset," [Online]. Available: https://www.kaggle.com/rounakbanik/pokemon (visited on 11/11/2020).

[2] Bulbapedia. (2020). "Statistic," [Online]. Available: https://bulbapedia.bulbagarden.net/wiki/Statistic (visited on 12/21/2020).

[3] ——, (2020). "Type," [Online]. Available: https://bulbapedia.bulbagarden.net/wiki/Type (visited on 12/21/2020).