Pokemon Analysis

Connor Bryson, Nathen Byford, Miguel Iglesias

10/12/2021

Introduction

Pokémon, a Japanese card and video game, revolves around a fantasy world where people fight each other with creatures they find and domesticate. Each creature, or Pokémon, can be characterized by certain attributes from health points (hp), to their typing which includes fire-Pokémon, water-Pokémon, and grass-Pokémon among others. Our dataset is a CSV that takes all the defined characteristics of the Pokémon and collects into one useable file. Though the dataset includes 49 variables, the ones we are using include: _____. One variable of interest, speed, is noteworthy because during the exploratory period of this project, we found that there was no relationship between speed and weight. This does not make intuitive sense, so we sought to identify if any other variable could correlate with speed.

Questions and Findings

Speed

We want to begin with looking into what factors effect the speed of the Pokemon.

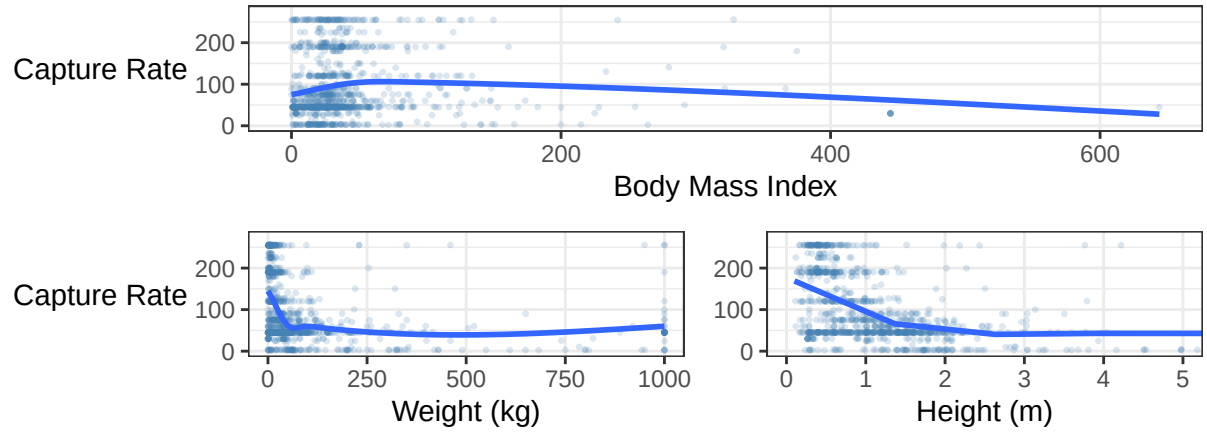
Capture Rate

We also wanted to look into how capture rate is effected by variables such as height, weight, gender, and legendary. Capture rate describes how easy it is to capture a Pokémon found in the wild. The lower the capture rate, the harder the Pokémon is to capture. Initially our belief was that the larger the Pokémon, the harder it would be to capture so we decided to compare capture rate to BMI, Height and Weight. We then wanted to see how certain categorical variables affected capture rate in the hopes of finding a pattern in how the game developers determined capture rate.

```
#Create a plot that shows the relationship between capture_rate and BMI
#Limit to not include a lot of empty space
BMI <- ggplot(pokemon_new, aes(BMI, capture_rate)) +
  geom_jitter(alpha = 0.2, color = "Steel Blue", size = .5) +
  geom_smooth(se = FALSE) +
  coord_cartesian(ylim = c(0,275)) +
  labs(x = "Body Mass Index", y = "Capture Rate")


#Create a plot that shows the relationship between capture_rate and Height
#Limit included to allow for best view of relationship
Height <- ggplot(pokemon_new, aes(height/10, capture_rate)) +
  geom_jitter(alpha = 0.2, color = "Steel Blue", size = .5) +
  geom_smooth(se = FALSE) +
```

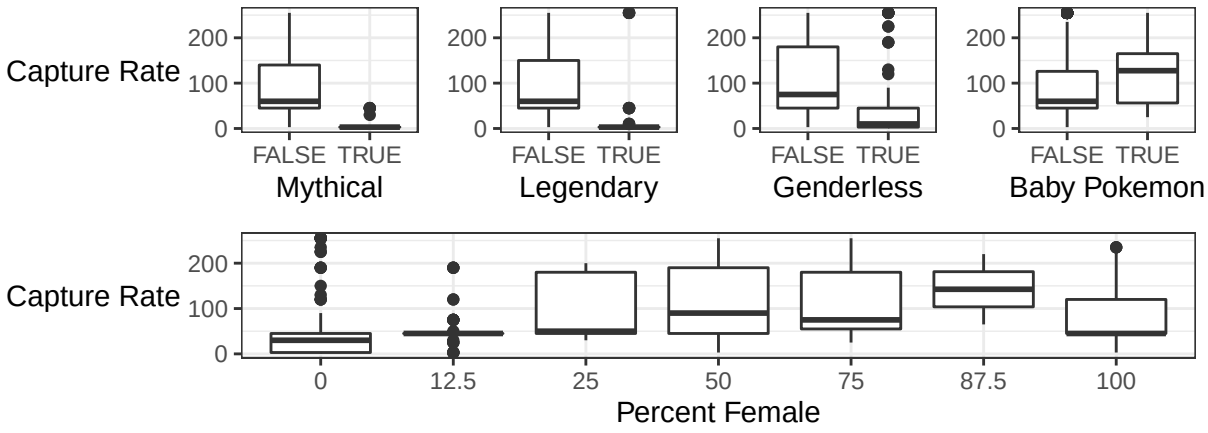
```
  coord_cartesian(xlim = c(0,5), ylim = c(0,275)) +
  labs(x = "Height (m)", y = NULL)

#Create a plot that shows the relationship between capture_rate and weight
#Limit included to allow for best view of relationship

Weight <- ggplot(pokemon_new, aes(weight/10, capture_rate)) +
  geom_jitter(alpha = 0.2, color = "Steel blue", size = .5) +
  geom_smooth(alpha = 0.5, se = FALSE) +
  coord_cartesian(ylim = c(0,275)) +
  labs(x = "Weight (kg)", y = "Capture Rate")


BMI / (Weight | Height)
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
p1 <- ggplot(pokemon_new) +
  geom_boxplot(aes(reorder(shape, capture_rate, FUN = median), capture_rate)) +
  labs(x = "Shape", y = "Capture Rate")

p2 <- ggplot(pokemon_new, aes(mythical, capture_rate)) +
  geom_boxplot() +
  labs(x = "Mythical", y = "Capture Rate")

p3 <- ggplot(pokemon_new, aes(legendary, capture_rate)) +
  geom_boxplot() +
  labs(x = "Legendary", y = NULL)

p4 <- ggplot(pokemon_new, aes(genderless, capture_rate)) +
  geom_boxplot() +
  labs(x = "Genderless", y = NULL)

p5 <- ggplot(pokemon_new, aes(baby_pokemon, capture_rate)) +
```

```
  geom_boxplot() +
  labs(x = "Baby Pokemon", y = NULL)

p6 <- ggplot(pokemon_new, aes(as.factor(female_rate*100), capture_rate)) +
  geom_boxplot() +
  labs(x = "Percent Female", y = "Capture Rate")

(p2 | p3 | p4| p5) / p6
```

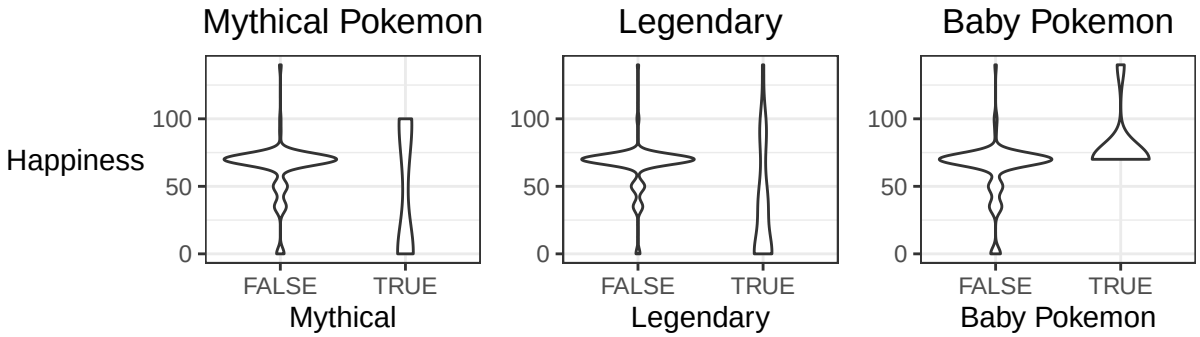


Happiness

Lastly we wanted to see what factors effected the happiness of the Pokemon. This variable was just something we thought would be fun to look into. Our main questions were what other variables effect the Pokemon's happiness. We began with looking at if the Pokemon is a legendary and if it's a mythical Pokemon.

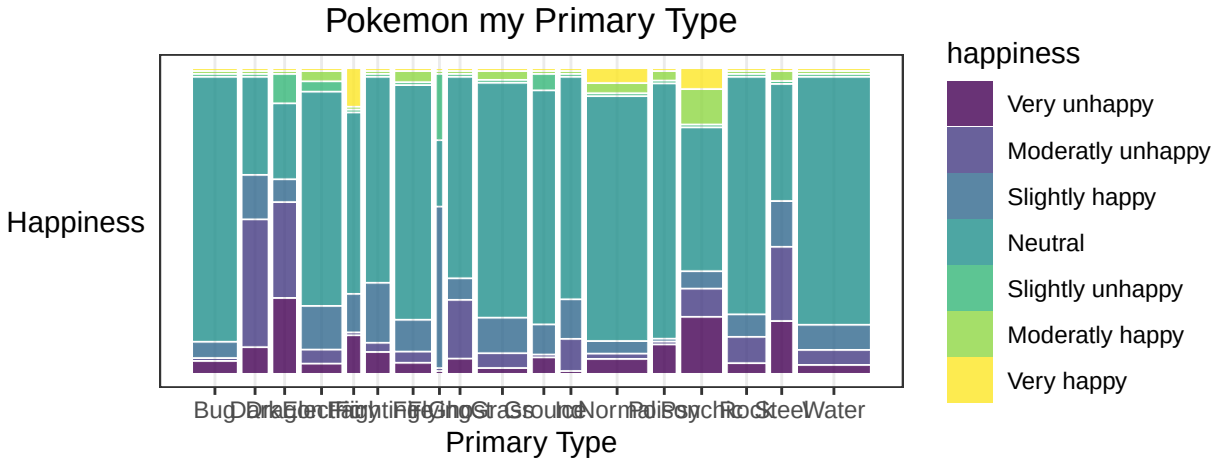
To get a grasp of what effects different variables had on the happiness of the Pokemon we took a look at how different characteristics of the Pokemon had an effect. Bellow are three violin plots looking at mythical, legendary, and baby Pokemon and comparing them to Pokemon without that attribute based on happiness. It is fairly obvious that mythical and legendary Pokemon tend to be less happy, I thinkk that could be that they are tougher pokemon so when they are captured they are not happy about it. These Pokemon like to have free reign and not be controlled by trainers. While the Baby Pokemon look to be as happy or more happy than most other pokemon.

```
v1 <- pokemon_new |> ggplot(aes(mythical, base_happiness)) +
  geom_violin() +
  labs(title = "Mythical Pokemon", x = "Mythical", y = "Happiness")

v2 <- pokemon_new |> ggplot(aes(legendary, base_happiness)) +
  geom_violin() +
  labs(title = "Legendary", x = "Legendary", y = NULL)

v3 <- pokemon_new |> ggplot(aes(baby_pokemon, base_happiness)) +
  geom_violin() +
  labs(title = "Baby Pokemon", x = "Baby Pokemon", y = NULL)

(v1 + v2 + v3)
```



We also looked at how the Pokemon primary type effects it's happiness. Looking at the mosaic plot below we can see that there are different distributions of happiness bases upon primary type. Some observations to be made is that the most unhappy pokemon types are dragon, dark, and steel. Also the majority of Pokemon appear to be neutral.

```
m1 <- pokemon_new |> ggplot() +
  geom_mosaic(aes(x = product(happiness, primary), fill = happiness)) +
  labs(title = "Pokemon my Primary Type", x = "Primary Type",
       y = "Happiness") +
  theme(
    panel.grid.major.y = element_blank(),
    axis.text.y = element_blank(),
    axis.ticks.y = element_blank()
  )

m1
```



We were also curious if the female rate effected happiness of the Pokemon. Converting the happiness variable to a likert scale we can easily see the amount of data points in each bin of female rate in the following plot. As seen in the plots before there is a smaller amount of Pokemon with a high female rate. Notice how it appears that depending of the female rate there are different levels of happiness that are observed. What stood out to us what that Pokemon with a female rate of 0.25 or 0.875 are only observed with neutral happiness and there are no very unhappy Pokemon with a rate of 1. There is no apparent correlation between happiness and female rate.

```
pokemon_new |> ggplot(aes(as.factor(female_rate), happiness)) +
  geom_bin2d(alpha = .9) +
  scale_fill_viridis_c() +
  labs(title = "Happiness of Pokemon by Female Rate", x = "Female Rate", y = "Happiness") +
  theme(panel.grid.major.x = element_blank())
```