Pokémon Analysis

Connor Bryson, Nathen Byford, Miguel Iglesias

10/12/2021

Introduction

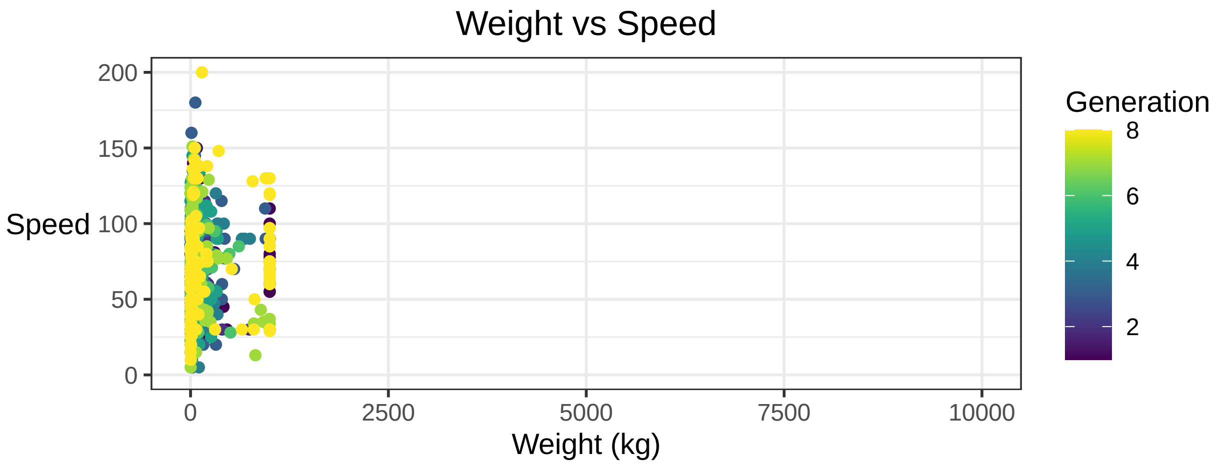
Pokémon, a Japanese card and video game, revolves around a fantasy world where people fight each other with creatures they find, capture, and then domesticate. Each creature, or Pokémon, can be characterized by certain attributes including type (fire, grass, poison etc.), identifiers (legendary vs non-legendary etc.), and skill points (health, attack, defense etc.). Within the world of Pokémon, this information is stored in what is known as the Pokedex. Our dataset is a CSV that takes all the attributes of every pokemon and collects into one usable, real-world version of the Pokedex. Though the dataset includes 49 variables, we primarily looked at speed, capture rate, and happiness. We choose these variables for a variety of reasons. Speed was chosen, because during the exploratory portion of the project we found that speed and weight had no significant correlation. So, I [Connor] decided to investigate if the developers of the game correlated speed to any other variables. The lack of relationship between weight and speed made us distrust the development of the game. It seemed as though the developers did not put much thought into the values they gave each Pokémon. Given that we were already looking at speed, I [Miguel] wanted to see if the capture rate had any predictors. I [Nate] decided to do the same but with the base happiness of the Pokémon. Each portion of the analysis will be discussed in more detail throughout the paper to give context to variables and analysis ideas.

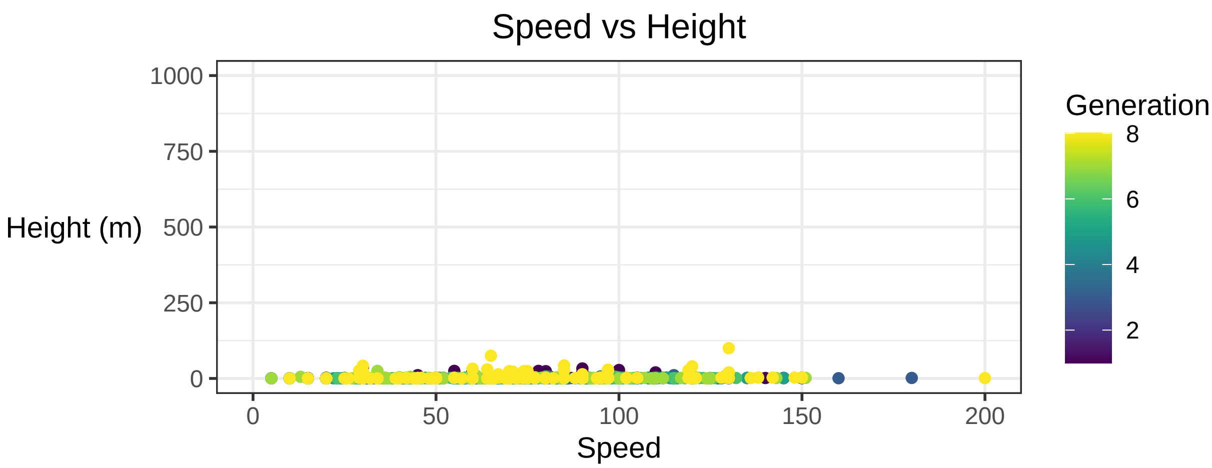
Figure 1: The pokedex entry for Pikachu. Includes info such as type, height, and weight

Questions and Findings

Speed

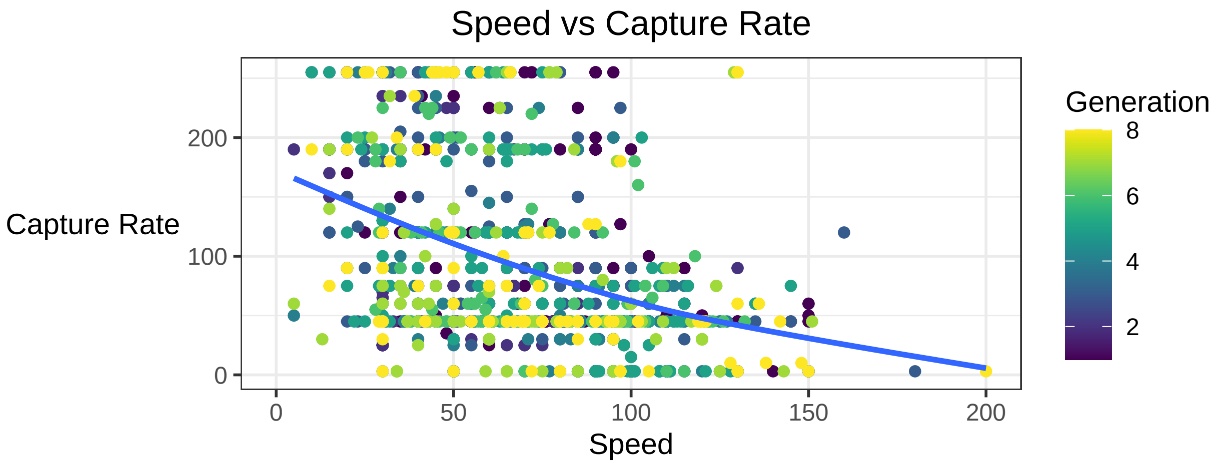
I [Connor] want to begin with looking into what factors effect the speed of the Pokémon. This included testing whether real life factors such as weight and height affected the speed of the Pokémon. At first we believed they would, but as seen by the graphs below we were wrong.





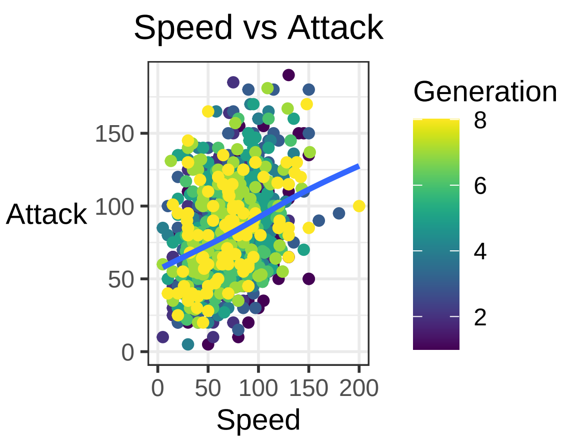
Looking at the graph for weight and speed, we noticed that there are outliers (on the right side of the graph) that do not make sense for speed. To see whether this was a mistake or not, we looked at the generation of each Pokémon and found that the majority of the outliers came from the most recent generation rather than the earlier generations when Pokémon's system of attributes were still in the works. In addition to weight, height did not have an affect on the speed of a Pokémon, thus defying all of our expectations of this data set.

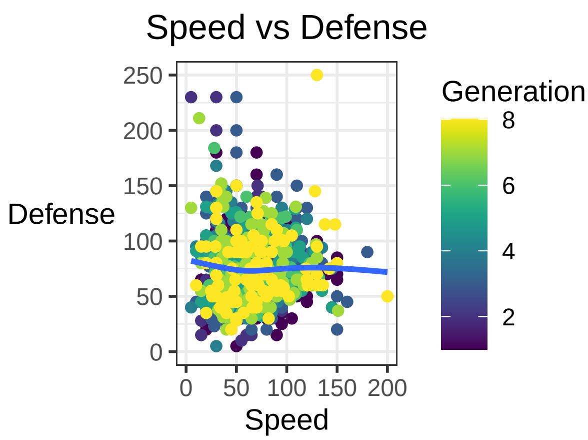
Afterwards, we wanted to see if the speed of a Pokémon affected its overall capture rate and whether it is harder to catch a Pokémon if it has a higher speed or not.



The plot shows a general negative trend as the speed increases, the capture rate decreases.

Finally with speed, since there was no realistic interpretation of speed for a given Pokémon, we decided to see if there was a relationship between a Pokémon's speed of attack and defense.

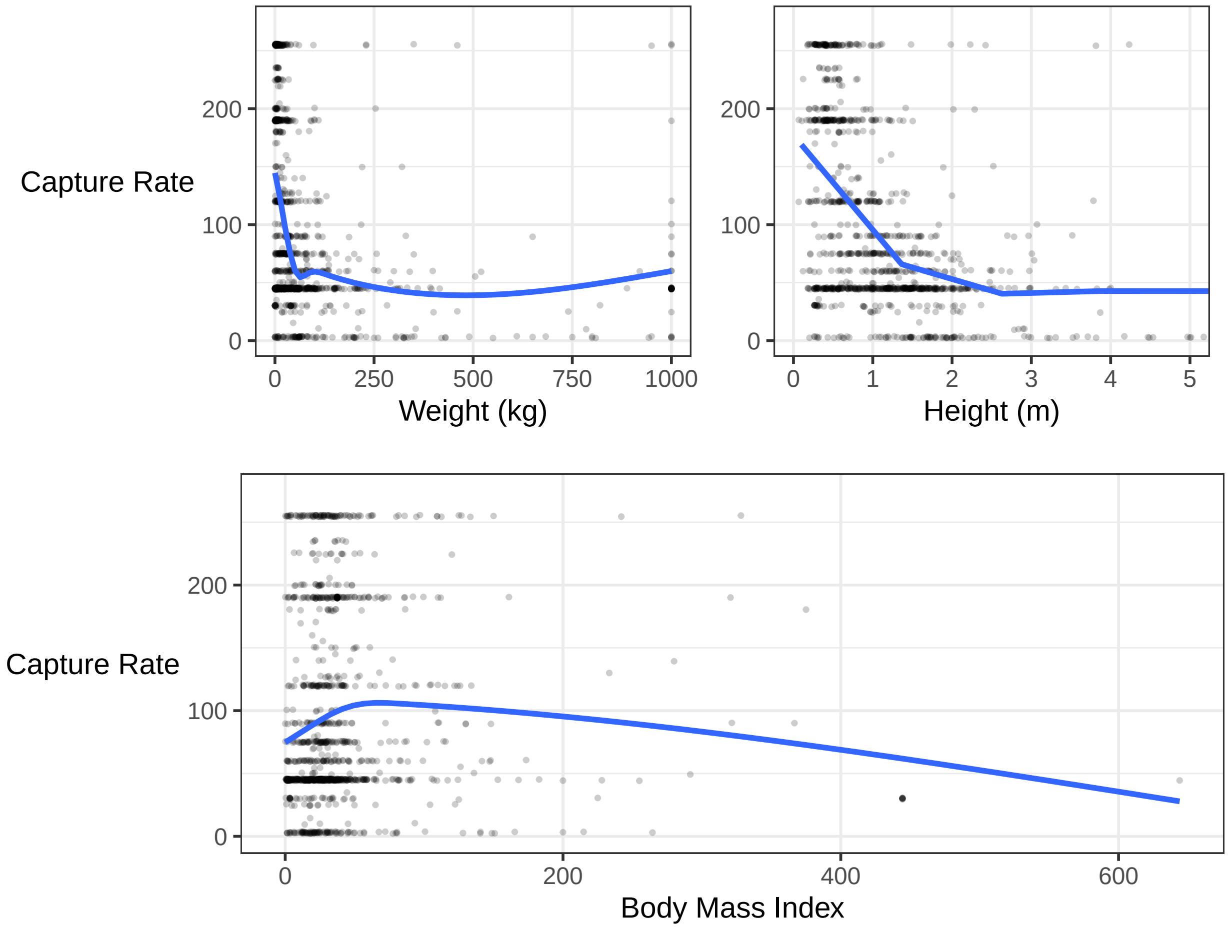




The results found that attack tends to have a slight positive correlation with speed while defense does not.

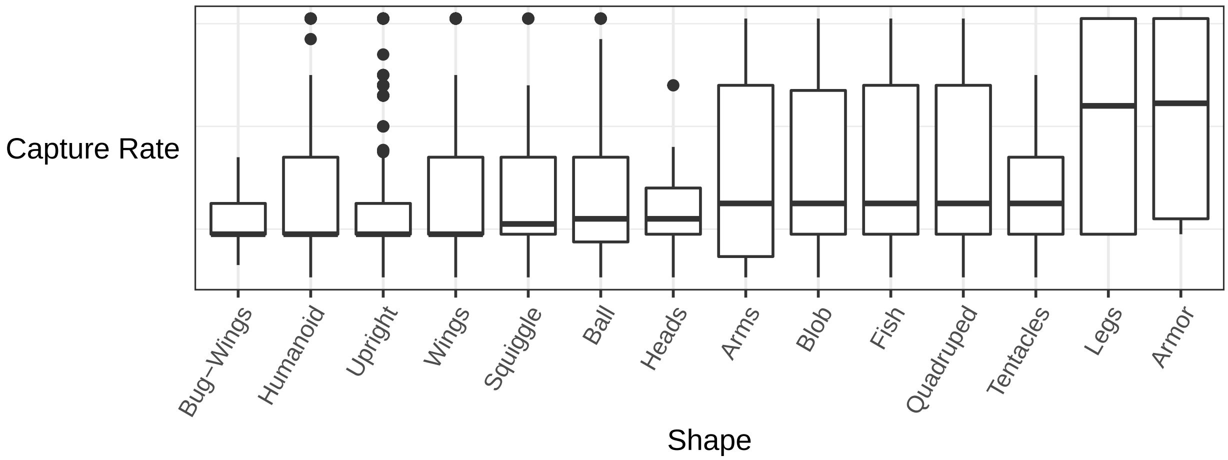
Capture Rate

Size versus Capture Rate As mentioned earlier, I [Miguel] wanted to test the game developers and see if anything predicted capture rate. 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. For more information on capture rate, a detailed description can be found at: https://bulbapedia.bulbagarden.net/wiki/Catch_rate. Of the variables in the dataset I first decided to look at the influence of height, weight, and then BMI. It made intuitive sense to me that bigger Pokémon would be harder to capture so I decided to test it out.



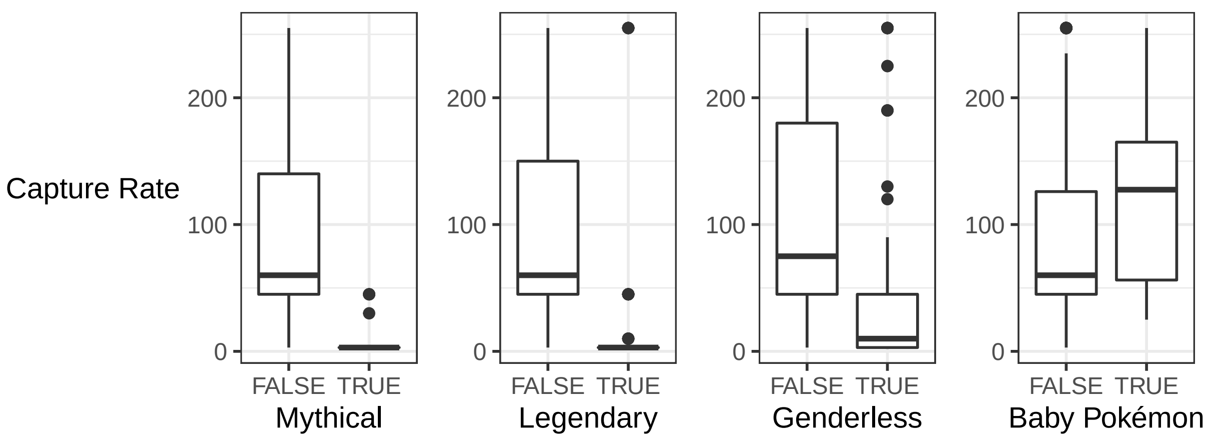
As can be seen, size does have an influence on the rate of capture. The taller and heavier Pokémon both have decreased rates of capture. Interestingly enough, however, when the two are put together as with BMI, the relationship goes away. To further understand the complex relationship between size and capture rate, more analysis would be required, however, the preliminary look suggests that size does have some effect on capture rate.

Other variables of interest After learning more about the relationship between size and capture rate, I decided to see how other variables may affect the capture rate. The first one I examined was the shape of the Pokémon. Shapes include armor, quadraped, upright and more.



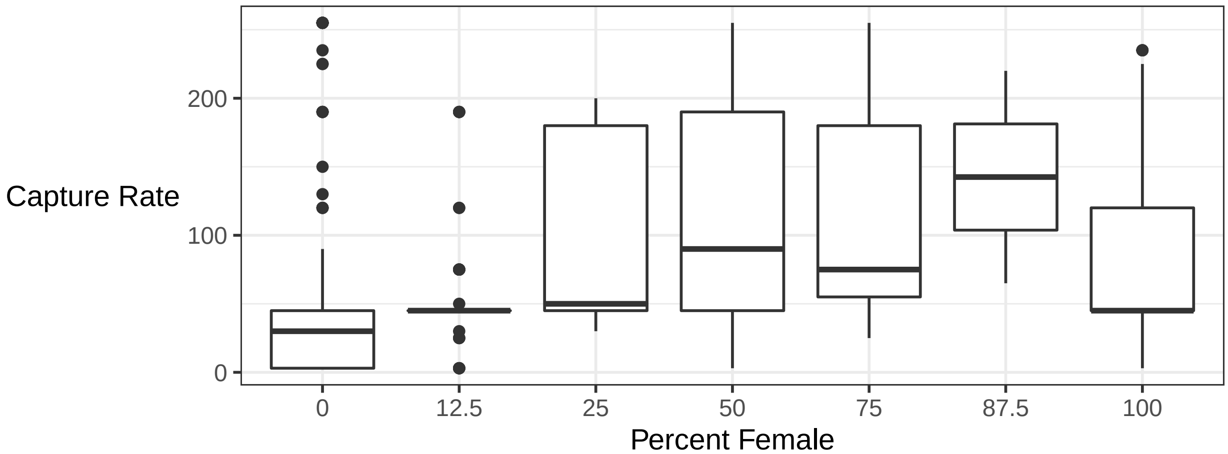
As seen in the plot above, shape has little influence on capture rate except for in the cases of Legs and Armor. What may be surprising, however, is that armored Pokémon are easier to capture than non-armored Pokémon.

Next, I looked at four of the boolean variables from the dataset. I wanted to see if how capture rate was influenced by Pokémon being Legendary, Mythical, genderless, and babies. One might expect mythical and legendary Pokémon to be harder to catch, babies easier to catch, and genderless Pokémon to have about equal rates to the gendered ones.



As expected, legendary and mythical Pokémon are more difficult to catch, and babies are easier to catch. What is surprsing, however, is that genderless Pokémon are also harder to catch. This could be due to a confounding variable, however, it is more likely that the game developers did not put that much thought into this. In fact, a preliminary look for confounding variables suggested that the genderless variable is not confounded by the legendary or mythical variable, both of which might have explained the increased difficulty of capture for genderless Pokémon.

Lastly, I looked at how the percentage that each Pokémon is female influences capture rate. Within the game, Pokémon can be bred to make new Pokémon. Due to the presence of Pokémon breeding, Pokémon gender must also exist and each Pokémon has a set probability of being female when found in the wild. Some Pokémon have a female rate of 12.5%, whereas others are as high as 100%. This does not mean a Pokémon can be 12.5% female and 87.5% male, merely that within the whole population, 12.5% of the Pokémon are female. Genderless Pokémon have a female rate of 0. For more information on the female rate, read: https://bulbapedia.bulbagarden.net/wiki/List_of_Pok%C3%A9mon_by_gender_ratio.

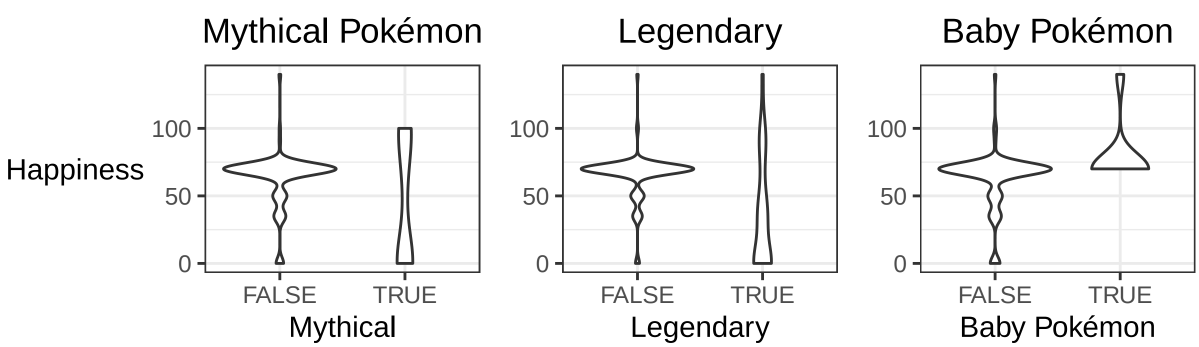


To avoid any confusion, it must be understood that within each Pokémon, the males and females have each capture rate. Though, as it turns out, the female rate does influence the capture rate. Generally speaking, the Pokémon with high female rates are easier to capture. Though the 0% female category may be lower due to the addition of the genderless Pokémon, the 12.5% and 25% catergories are also low, especially compared to the 87.5% category. Why male-dominant Pokémon are harder to capture than female-dominant Pokémon is a question for another day.

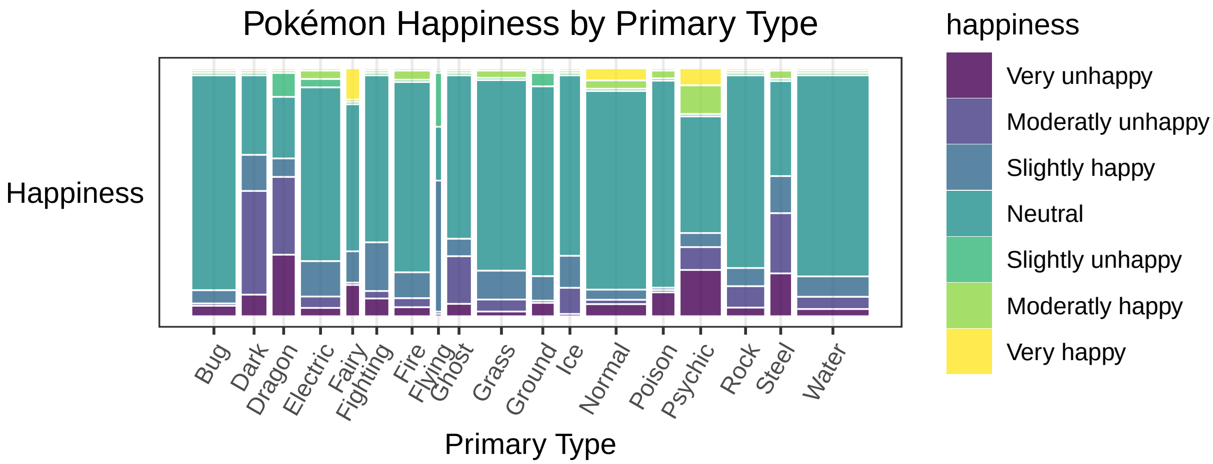
Happiness

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

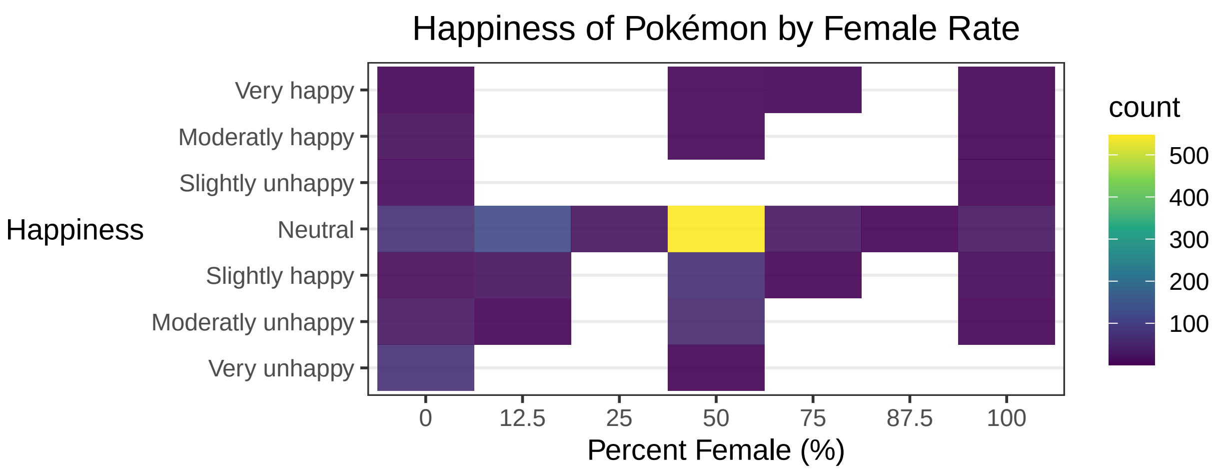
Bellow are three violin plots looking at mythical, legendary, and baby Pokémon, comparing them to Pokémon without that attribute based on happiness. It is fairly obvious that mythical and legendary Pokémon tend to be less happy, we think this could be because they are tougher Pokémon so when they are captured they are not happy about it. Maybe these Pokémon like to have free reign and not be controlled by trainers. While the Baby Pokémon look to be as happy or more happy than non baby Pokémon.



We also looked at how the pokémon 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 pokémon appear to be neutral. It can also be noted that fairy and psychic pokémon seem to be the happiest.



Another variable we wanted to look at was how the female percentage of each pokémon effected happiness of the Pokémon. 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 Pokémon 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 Pokémon with a female rate of 0.25 or 0.875 are only observed with neutral happiness and there are no very unhappy Pokémon with a rate of 1. There is no apparent correlation between happiness and female rate.



Conclusion

For a game as popular as Pokémon, one might expect that careful planning has dictated every decision; every statistic for every Pokémon being thoughtfully decided with careful analysis. After all, if a game makes no sense, it would be unplayable. Given the size of the Nintendo company, there should be no excuse behind what our group has discovered; that whoever designs each Pokémon has no idea why they make the decisions that they do. This is especially true today, as shown by the plot comparing speed and weight by each generation. Whether weight or height play factors for speed is ignored. This can also be seen when looking at capture rate; variables like height, weight, legendary, status as a baby each can predict capture rate. Other variables like BMI and shape do not, however. Now, it is true that some variables have intuitive correlations, such as height and capture rate. But what is truly baffling is that there seems to be no pattern behind which relationships make sense and which ones do not. We also see that Pokémon happiness seems to be related to type and certain attributes like weather they are a legendary or a baby. Why is speed correlated to attack values, but not defense? Why does shape have little influence on capture rate, except in the cases of leg and armor?

In a dataset rife with unexplainable associations, our group did find one variable that influences numerous other variables. That variable is the female rate. Female rate influenced both the capture rate, and the happiness. Though not included in this analysis, during our exploratory portion of our analysis, we found other variables that were influenced by female rate, especially the primary color. The reason that female rate is such a strong predictor may make sense from a sociological real-world perspective, however, the reasoning behind our findings within the Pokémon world remains a mystery.

Overall, our analysis unearthed more questions than answers, and much more about the world of Pokémon needs to be investigated. More analysis into possible confounding variables may help explain some of our findings such as the lower capture rate for genderless Pokémon or why the latest generation of pokemon is the least happy. Future research topics could also include searching for actual interviews or statements by the developers regarding the decision making process when creating new Pokémon. Should we ever meet some of the people responsible for designing the game, we would have numerous questions to ask but, in the meantime, we hope that the logic, or lack thereof, behind Pokémon creation can be better understood by our work.

Appendix

Data Source

https://www.kaggle.com/kylekohnen/complete-pokemon-data-set

Background Information Sources

Catch Rate Info: https://bulbapedia.bulbagarden.net/wiki/Catch_rate

Female Rate Info: https://bulbapedia.bulbagarden.net/wiki/List_of_Pok%C3%A9mon_by_gender_ratio

Pokedex Picture: https://bulbapedia.bulbagarden.net/wiki/Pok%C3%A9dex

For more information on Pokémon in general, good resources include:

The fan-run wiki: https://bulbapedia.bulbagarden.net/wiki/Main_Page The official website: https://www.pokemon.com/us/ The actual show: https://www.netflix.com/watch/81198528?source=35&trackId=254743534

Code

```
# read in data.
pokemon <- read_csv(".\\data\\pokemon.csv")

# make function to factorize happiness
factorize <- function(x) {
  factor(case_when(
            x == (0) ~ "Very unhappy",
            x == (35) ~ "Moderatly unhappy",
            x == (50) ~ "Slightly happy",
            x == (70) ~ "Neutral",
            x == (90) ~ "Slightly unhappy",
            x == (100) ~ "Moderatly happy",
            x == (140) ~ "Very happy"
```

```
            ),
          levels = c(
            "Very unhappy",
            "Moderatly unhappy",
            "Slightly happy",
            "Neutral",
            "Slightly unhappy",
            "Moderatly happy",
            "Very happy"
            ),
          ordered = TRUE
          )
}

# New pokemon dataset to use, adds BMI and happiness. BMI was filtered to
#remove one outlier with a BMI of 100000. The next highest BMI was 644.
pokemon_new <-  pokemon %>%
  separate(typing, into = c("primary", "secondary")) %>%
  mutate(BMI = (weight / 10) / ((height / 10)^2)) %>%
  filter(BMI <= 50000) %>%
  mutate(happiness = factorize(factor(base_happiness, ordered = TRUE)))

#Weight vs Speed (organized by Generation)
spdvswgt <- ggplot(pokemon, aes(x =weight, y =speed, color= gen_introduced)) +
  geom_point()+
  labs(title="Weight vs Speed", x= "Weight", y="Speed")+
  xlim(c(0, max(pokemon$weight))) +
  ylim(c(0, max(pokemon$speed)))+
  viridis::scale_color_viridis()


#Spvs Height (organized by Generation)
spdvshgt <- ggplot(pokemon, aes(x =speed, y =height, color= gen_introduced)) +
  geom_point()+
  labs(title="Speed vs Height", x= "Speed", y="Height")+
  xlim(c(0, max(pokemon$speed))) +
  ylim(c(0, max(pokemon$height)))+
  viridis::scale_color_viridis()+
  theme(plot.title = element_text(hjust = 0.5))

spdvswgt
spdvshgt

#Speed vs Capture Rate
ggplot(pokemon, aes(x =speed, y =capture_rate,  color= gen_introduced)) +
  geom_point()+
  labs(title="Speed vs Capture Rate", x= "Speed", y="Capture Rate")+
  xlim(c(0, max(pokemon$speed))) +
  ylim(c(0, max(pokemon$capture_rate)))+
  geom_smooth(se=FALSE)+
  viridis::scale_color_viridis()+
  theme(plot.title = element_text(hjust = 0.5))
```

```r
#Speed vs Attack
spdvsatk <-ggplot(pokemon, aes(x =speed, y =attack,  color= gen_introduced)) +
  geom_point()+
  labs(title="Speed vs Attack", x= "Speed", y="Attack")+
  xlim(c(0, max(pokemon$speed))) +
  ylim(c(0, max(pokemon$attack)))+
  geom_smooth()+
  viridis::scale_color_viridis()+
  theme(plot.title = element_text(hjust = 0.5))

#Speed vs Defense
spdvsdef <-ggplot(pokemon, aes(x =speed, y =defense,  color= gen_introduced)) +
  geom_point()+
  labs(title="Speed vs Defense", x= "Speed", y="Defense")+
  xlim(c(0, max(pokemon$speed))) +
  ylim(c(0, max(pokemon$defense)))+
  geom_smooth()+
  viridis::scale_color_viridis()+
  theme(plot.title = element_text(hjust = 0.5))

spdvsatk+spdvsdef


#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, 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, 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, size = .5) +
  geom_smooth(alpha = 0.5, se = FALSE) +
  coord_cartesian(ylim = c(0,275)) +
  labs(x = "Weight (kg)", y = "Capture Rate")


(Weight | Height)
BMI
```

```
#Boxplot comparing the capture rate for different Pokémon shapes
p1 <- ggplot(pokemon_new) +
  geom_boxplot(aes(reorder(shape, capture_rate, FUN = median), capture_rate)) +
  labs(x = "Shape", y = "Capture Rate")

p1


#Boxplots comparing the capture rate for different boolean variables

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 Pokémon", y = NULL)

p2 | p3 | p4 |p5


#Boxplot to examine difference between capture rate and female rate

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

p6

v1 <- pokemon_new |> ggplot(aes(mythical, base_happiness)) +
  geom_violin() +
  labs(title = "Mythical Pokémon", 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 Pokémon", x = "Baby Pokémon", y = NULL)

(v1 + v2 + v3)

m1 <- pokemon_new |> ggplot() +
  geom_mosaic(aes(x = product(happiness, primary), fill = happiness)) +
```

```r
  labs(title = "Pokémon Happiness by Primary Type", x = "Primary Type",
       y = "Happiness") +
  theme(
    panel.grid.major.y = element_blank(),
    axis.text.y = element_blank(),
    axis.ticks.y = element_blank(),
    axis.text.x.bottom = element_text(angle = 60, vjust = 1, hjust = 1)
  )

m1


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