Summary

Question # 1 How many Pokemon were introduced per generation?
To visualize our findings, we created a bar chart that separated each bar by generations 1 - 8. Each bar shows the total amount of pokemon added to that generation. The bar totals are as follows: 151, 100, 135, 107, 156, 72, 88, 96. We were interested in the findings. The odd generations had larger outputs of Pokemon until Gen 7. We also find that generations 6, 7, 8 have the smallest amount of created pokemon. We come to the question of maybe the company is running out of ideas for new Pokemon. This is a limitation of our data.

Question #2 How are the Pokemon attributes distributed across types?
When showing out findings, we created a scatter plot to include each pokemon separated by stat and type. Each type has it's own color to visualize where the Pokemon falls for each battle stat. When looking at the data/chart we can find the outliers, both above and below where the large majority of Pokemon fall. This leads us to another question about what is common between these outliers. We find that the outliers above the average are legendary/mythical Pokemon.

Question #3 How are the attributes distributed?
To answer this question, we created a box and whisker plots that graphs the attributes such as Hit Points and Speed against the numerical values the pokemon have in each category. Plotting each box and whisker plot on the same set of axes ensures that each graph can be easily compared.

Our findings include that the median values for all of the stats are close to each other. We can see that there are a good number of outliers above the box plots which leads into our next question, however I want to discuss the single lower outliner. Seeing this outliner, it sparked our attention to ask "Who's that Pokemon?!" It's Shedinja with an HP value of 1.
Shedinja is the only Pokémon to have a stat not calculated by the standard formula; if it were, Shedinja would have at least 11 HP even at level 1.

Looking at the upper outliers, we asked if there is a commonality amongst these pokemon. We found a majority of these are not "normal". And normal here does not refer to types like described in slide 5, but rather to legendary status. Next we looked at distribution of pokemon on their legendary status.

Question #4 How are Pokemon distributed by status?
The options for this property are Normal, Mythical, Sub Legendary, and Legendary. Because of these limited categories, we felt that a pie chart would work well to highlight how small the non-normal population is.

Some examples of Mythical pokemon are Mew (Gen 1), Celebi (Gen2), and Jirachi (gen3) Mythical can loving be summarized as the cute powerful babies. Mythical pokemon makeup 2.9 % of pokemon in our data set.

Sub Legendary are Pokemons such as Moltres, Articuno, and Zapdos featured on the slide. Notably these are Team logos for the mobile game, Pokemon Go! Sub Legendary pokemons are nice sets of pokemon with a theme. Sub Legendary makes up 5.3%

Legendary pokemon are usually the game cover pokemon, for example Gen 3 games are titles Ruby, Sapphire and Emerald with the 3 legendaries featuring Groudon, Kyogre, and Rayquaza respectively. Legendary make 3.8%

Changing gears, we decided to look into other attributes of Pokemon, such as height and weight.

Question #5 How does the height and weight of pokemon correlate?
To start our examination into the height/weight relationship, we created a scatter plot of height vs. weight. We chose to do so so that we could see the individual pokemon. What we can see is that a majority of pokemon are "small" boys tightly fit between 0-2 meters and 0-200 kg or 0-6.5 feet and 0-440 lbs.

What peaked our interest in this graph was some of the outliers, notable the overall tallest pokemon is Waillord, the literally blue whale pokemon. The three heaviest are Celesteela, Cosmoem, Primal Groudon. Celesteela is covered more in a later chart. We were definitely intrigued in Cosmoem, Protostar Pokemon, featured here on the slide. I enjoy seeing pokemon based on real like bugs and animals, so this one based on a protostar.

Despite clocking in at 999.9 kg or 2204 lbs, the anime features him being easily picked up.

This chart and linear regression inspired our next question looking at the height of pokemon per generation.

Question #6 What is the tallest Pokemon per generation per Type 1?
Here we have sorted the dataset to separate each generation by type to determine the tallest Pokémon per generation. For generation 1 originally the tallest Pokémon was Onix for Rock type Pokémon till Alolan Exeggutor was created that changed grass type to being tallest for gen 1. Based on our dataset we have included images of the tallest Pokémon for each generation.

Gen 1: Onix(Rock) originally till Alolan Exeggutor (Grass)
Gen 2:Steelix(steel)
Gen 3:Wailord(water)
Gen 4: Giratina Origin Forme(ghost)
Gen 5: white kyurem(dragon)
Gen 6: Hoopa Unbound(psychic)
Gen 7: celestella(steel)
Gen 8: Sandaconda(ground)

Question #7 Is there a relationship between a Pokemon's type vs. the sum of its stats?
Here we have a scatter plot on top and a box plot on the bottom that graph the total points of each pokemon by their first type which are labeled on the x axis. (grass type, fire type, etc)

What we based total points off of is the sum of all their attributes AKA stats which includes (hit points, attack, defense, speed) all that stuff that was defined in our Stats 101 slide

So the scatter plot is pretty much self explanatory. Each little dot is a pokemon and it's total points. You can easily see the outliers and somewhat of an average point range where the dots are bunched up.

The box plot is the same information as above just a different visual. The bars are graphed at the average and the little black line in the middle of each bar which is called an "error bar" is showing us the outliers. It's pretty subtle
in most of them due to there being a pretty even number of low vs high outliers. So the black line inside of the bar is the low outliers which would be the pokemon 250 points and below and the line past the bar is the high outliers which would be pokemon above like 550 points.

Here in the next slide, we did the same thing, but for each pokemon's second type. Their 2nd type's total points are a lot more scattered. Now there are way more outliers. Again, most of of them have pretty even high vs low outliers, but if we look at the "Fighting type" pokemon, you can see very obviously that the line inside is shorter than the line outside which means there are less low outliers than high outliers which is proven to be true in the scatter plot. There are no fighting type pokemon below 300 total points. So I feel like a bar plot would be very beneficial if your data has extreme differences.

Sources

Pokemon with Stats
- Served as inspiration of what we can do with Pokemon data

Pokemon_Data
- Served as inspiration of what we can do with Pokemon data

Complete Pokemon Dataset (Updated 16.04.21)
- The main data set that we used

Bulbapedia
- Used for cleaning the data, filling in missing details. Also where we got our images