3300 Project 1
Jill Wu, Stefan Joseph, Christian Llamas

Pokemon Go Visualization

For our visualization, we took a Pokemon Go visualization from Kaggle. (https://www.kaggle.com/semioniy/predictemall)

This dataset is a record of Pokemon caught over the span of 7 days, from 9/2/2016 to 9/8/2016. The dataset includes basic information about the Pokemon, including the Pokemon ID, the location and time of where and when the Pokemon was caught (including latitude, longitude, and city). Additionally, there were other variables such as how close the Pokemon was to a gym or Pokestop and the type of area the Pokemon was found in (rural, midurban, suburban, and urban).

The dataset was extremely large (300,000 data points, around 0.5GB data file), so we decided to filter down the data so that it would be both easier to work with and so that we could focus on particular cities. We also took out data that we didn't find helpful, such as the dummy day columns, class (same data as PokemonId), appeared day of week, and co-occurence columns. We decided to create three different visualizations:

1) A frequency graph of how often the Pokémon showed up, mapped over Los Angeles, New York, and Chicago
2) A pair of bar graphs that compare the number of times rare Pokémon as well as common Pokémon spawned in urban and rural areas
3) A group of US maps that showcase the amount of encounters people had with four original pokemon: Pikachu, Bulbasaur, Squirtle, and Charmander.

While selecting our data, we tried to use variables that were a little more interesting and gave a little insight into the Pokémon that appeared in the game. For example, the second visualization squarely juxtaposes rural and urban areas in order to compare the gameplay experience in each region. Players often complain that spawns in rural areas are vastly inferior to those in urban areas, so we thought this would be an intriguing set of metrics to study.

For the first visualization, we decided that a circular approach was the best way to display the data, because it mapped hours in a day to Pokémon count. We categorized the Pokémon counts using a 24-hour scale, and created a bar graph organized in a circular shape. The top of the circle would indicate noon (hour 12), and the bottom of the circle would indicate midnight (hour 0). We created bars extending from the circle, with varying heights to indicate the average count of Pokémon appearances for that time of day. In order to allow the viewer to understand the heights of the bars, we also created varying concurrent circles that acted as axis ticks; each axis tick would increment by 300. There were also labels created for each bar, to indicate the hour that the bar matched to.

In order to allow the viewer to easily understand day and night, we created two arcs within the graph, on the inner side of the circle. The yellow arc indicated daytime (average sunrise time to average sunset time), and the blue arc indicated nighttime (average sunset time to average sunrise time). These two colors typically indicate day and night (or are related to the colors of the sun and moon), so the viewer can easily differentiate between the two. Additionally, we added the images of the sun and moon that are featured in the Pokémon universe.

For the second visualization, we went with bar graphs in order to most clearly contrast the number of spawns for individual Pokémon between urban and rural areas. The blue bars represent the urban count of spawns for a given Pokémon while the green bars represent the same for Pokémon that spawned in rural areas. Overlapping bars of opposite colors correspond to the same Pokémon. We further divided the data by creating one graph consisting of relatively common Pokémon and another featuring more scarce Pokémon. The rarity of Pokémon was determined by calculating the most and least frequently occuring Pokémon in the narrowed dataset. Note that we ignored Pokémon that spawned less than 100 times in order to make the proportions between rural and urban counts more reflective of the data. Also note that "urban" and "rural" are determined by our dataset by using the population density at the point where the spawn occurred. For this visualization, we also narrowed our dataset to three major cities: New York, Los Angeles, and Chicago in order to make the data easier to digest for our code.

We then chose to focus on particular pokemon. We decided that it would be best to focus on the four classic pokemon from the series: Bulbasaur, Charmander, Squirtle, and Pikachu. We then filtered the data and took out only the entries that were in the United States and that involved only those four pokemon with the following pokemonId: 1, 4, 7, and 25 respectively. We did this in order to significantly reduce the data size in order to make it fast and efficient in loading and analyzing the data. We then acquired US json data and pathed it to make a map. We used this map to plot circles in the US cities that were in the data set and scaled them using the range zero to the max amount of that particular pokemon and decided on a range of radii for all the data sets that we determined to be visually appealing. We then grouped all the maps together so they can be seen together and one can see the differences of encounters with certain pokemon given a particular city. It is important to note that I deleted 3 US cities from the data set because they had less than 2 encounters of any of the type of pokemon we were focusing on. This lack of data wouldn't be beneficial in viewing (a speck on the map) and thus, we decided to filter it out of our data. We decided to give each pokemon its own map so that way other pokemon (circles) would be interfering with what the data was trying to portray. We colored each circle according to the pokemon's color: Squirtle-blue, Pikachu-yellow, Charmander-red, and Bulbasaur-green. It was also important to include a legend for each map since the range varies between each pokemon. For example, there was only 301 pikachu encounters in the US whereas there 575 squirtle encounters. We used the max of a particular city and pokemon as our max for our range when scaling the data to radii.

The Story

The visualization gives a general view of different statistics generated in the mobile game Pokemon Go. Players find Pokemon that “spawn” in the game by traveling to real-life locations; Pokemon that spawn are generally part of the first generation of the Pokemon Universe.

The dataset includes information about Pokemon that have “spawned” over the course of a week, in various states around the world. The first visualization gives us a look at how often Pokemon appear at certain times of day. From the graph, we can see that Pokemon tend to appear more in nighttime hours rather than daytime hours, with the most Pokemon appearing at 2am, and the least Pokemon appearing at 9 and 10am. Additionally, we can see a gradual increase of the amount of Pokemon appearances as the hours approach into the nighttime, and then a decrease as the hours approach into the daytime.

The second visualization suggests that urban areas have more plentiful spawns for Pokémon more commonly found in the game (see the graph labelled “Common Spawns”.) That said, the number of rural spawns is relatively close to the number of urban spawns in this graph. This is not particularly surprising; the number of Pokémon spawning in rural areas compare to urban ones was never the greatest point of contention amongst players. Players were much more concerned with the number of rare spawns in each area. The visualization responds to this qualm in the “Rare Spawns” graph. To our surprise, it shows a balanced amount of activity: though certain rare Pokémon do spawn more often in one area compared to the other, this happens twice for each area. The only outlier is the leftmost pair of bars that are nearly equal. We expected to see much more disparity here, with urban areas having many more rare spawns than rural ones. It is possible that our minimum stipulation of 100 spawns may have affected this outcome. That said, we still do not feel that Pokémon with fewer data points would have been worth considering in this minimized dataset.

The third visualization is able to show different trends in the United States. First and foremost, it shows that New York City was always the top city in terms of having the most encounters of any pokemon. However, one might think that Los Angeles would be in second place for every pokemon since it is a large city with a large population. However, it can be seen in the graphs that Chicago had more encounters with Bulbasaur than L.A. It is also interesting to see the data amongst the same city. For example, Phoenix had much less Bulbasaur encounters than Squirtle.

Citations

Visualization 1 is modeled off of the radial bar graph:
https://bl.ocks.org/mbostock/6fead6d1378d6df5ae77bb6a719afcb2
Additionally, the radial scale from
https://gist.github.com/mbostock/6fead6d1378d6df5ae77bb6a719afcb2#file-d3-scale-radial-js was used to create the bar graph.
Sun Picture:
https://vignette.wikia.nocookie.net/fantendo/images/e/e0/Pokemon_Sun_Icon.png/revision/latest?cb=20170624183657
Moon Picture: https://alexalan.deviantart.com/art/POKEMON-MOON-SYMBOL-593895361
Pokeball Picture:
https://www.flaticon.com/free-icon/pokeball_361998#term=pokeball&page=1&position=16
Pidgey Picture:
https://bulbapedia.bulbagarden.net/wiki/Pidgey_(Pok%C3%A9mon)
Rattata Picture:
https://bulbapedia.bulbagarden.net/wiki/Rattata_(Pok%C3%A9mon)
Weedle Picture:
https://bulbapedia.bulbagarden.net/wiki/Weedle_(Pok%C3%A9mon)
Spearow Picture:
https://bulbapedia.bulbagarden.net/wiki/Spearow_(Pok%C3%A9mon)
Eevee Picture:
https://bulbapedia.bulbagarden.net/wiki/Eevee_(Pok%C3%A9mon)
Pidgeot Picture:
https://bulbapedia.bulbagarden.net/wiki/Pidgeot_(Pok%C3%A9mon)
Fearow Picture:
https://bulbapedia.bulbagarden.net/wiki/Fearow_(Pok%C3%A9mon)
Koffing Picture:
https://bulbapedia.bulbagarden.net/wiki/Koffing_(Pok%C3%A9mon)
Seel Picture:
https://bulbapedia.bulbagarden.net/wiki/Seel_(Pok%C3%A9mon)
Metapod Picture:
https://bulbapedia.bulbagarden.net/wiki/Metapod_(Pok%C3%A9mon)

Visualization 3: Uses code from lecture: Prof Minmo, Lecture Feb 16th, 2018
Also, this example helped significantly in understanding how to read geoJson's
http://bl.ocks.org/michellechandra/0b2ce4923dc9b5809922
As well as provided the framework to construct the Legends on the maps in accordance with this Modified Legend Code from Mike Bostock: http://bl.ocks.org/mbostock/3888852
As well as the json file of the United States https://d3js.org/us-10m.v1.json