

# Data Project on Pokemon

DATA 110

James Pham

# The Dataset

## What is the Pokemon dataset?

The Pokemon dataset includes columns of the Pokemon's name, Trainer's region, Pokemon's region, Pokemon's level, Pokemon's type(s), and more.

I chose this because I'm somewhat familiar with Pokemon and Pokemon games and was interested in what I could find with this dataset.

Date	Time	Pokemon	Trainer Region	Trainer Subregion	Pokemon Region	Level	Level Met	Gender	Type1	Type2	Nature	Pokeball	Held Item	Perfect IVs
12/13/2016	17:28	Oricorio	South Korea			13	10	F	Electric	Flying	Sassy	Poke	F	0
12/13/2016	17:30	Zubat	United States	Texas	GER	8	8	M	Poison	Flying	Hardy	Poke	F	1
12/13/2016	17:31	Carbink	United States	Oklahoma		10	10	N	Rock	Fairy	Relaxed	Poke	F	0
12/13/2016	17:33	Klefki	United States	Connecticut		29	29	M	Steel	Fairy	Jolly	Quick	F	0
12/13/2016	17:34	Luvdisc	United States			16	16	M	Water		Naughty	Quick	F	0
12/13/2016	17:35	Roggenrola	United Kingdom		SPA	10	10	M	Rock		Modest	Poke	F	1
12/13/2016	17:36	Salandit	United States	California		19	19	M	Poison	Fire	Sassy	Quick	F	0
12/13/2016	17:38	Yungoos	United States	Indiana		7	7	M	Normal		Quiet	Great	F	0
12/13/2016	17:39	Wimpod	United States	New York		1	1	F	Bug	Water	Impish	Quick	F	3

# The Process (EDA)

## How did I approach this dataset?

First, I needed to look at each column and see if there were going to be any trends or if it was random.

To do this, I had to compare the different elements in the dataset.

Then I made some questions to guide this process.

# The Process (continued)

## Constraints

While trying to understand relationships between the elements in the dataset, I noticed that this dataset has majority of object values and less integer values.

This made it difficult for me to make graphs the way this class has taught me so far.

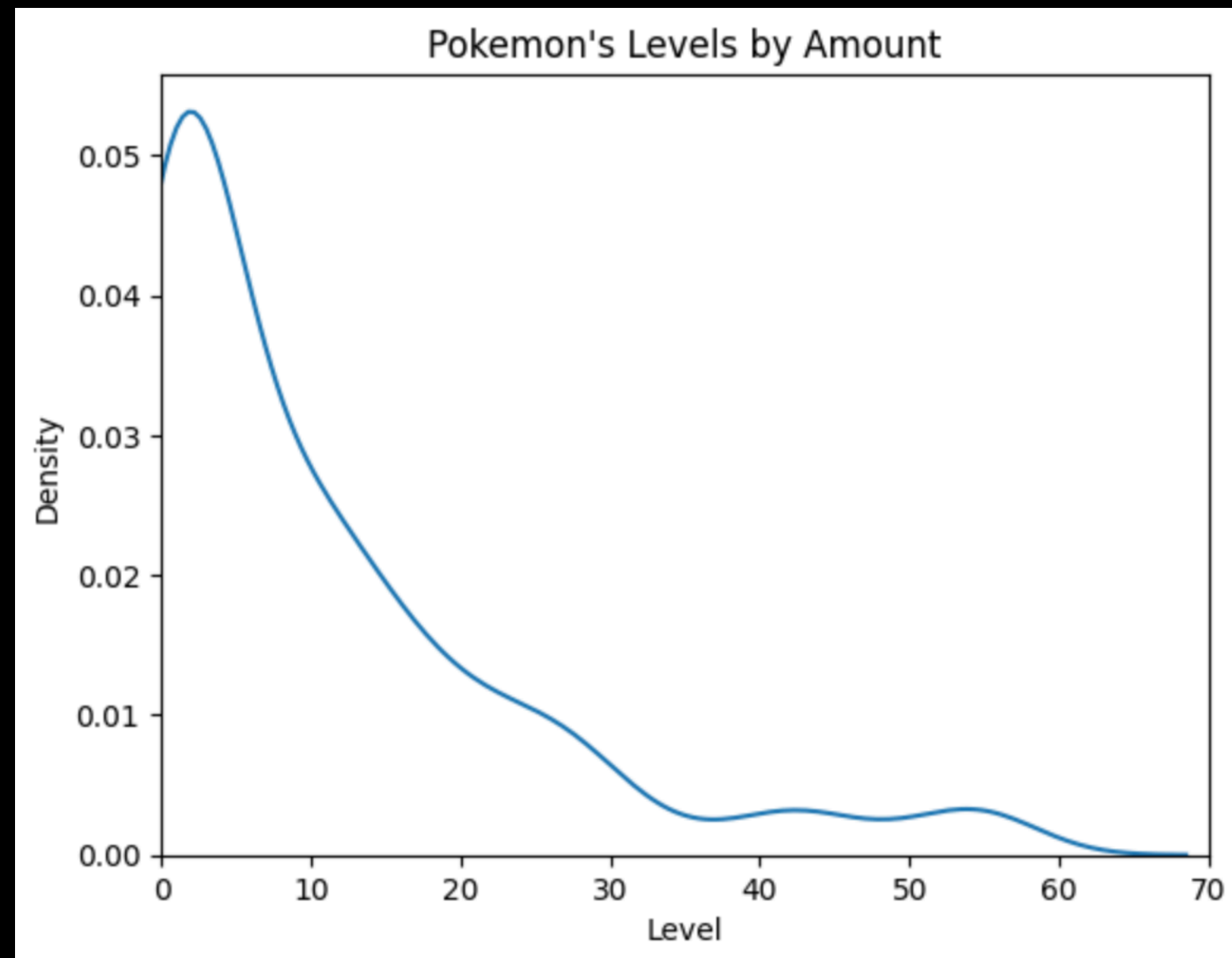
There are only 3 columns that have integer values as seen below:

pokemon														
Date	Time	Pokemon	Trainer Region	Trainer Subregion	Pokemon Region	Level	Level Met	Gender	Type1	Type2	Nature	Pokeball	Held Item	Perfect IVs
12/13/2016	17:28	Oricorio	South Korea			13	10	F	Electric	Flying	Sassy	Poke	F	0

# The Process (continued)

## Observations

Of the 501 Pokemon, the majority of them were between levels 1-10 (as seen below)

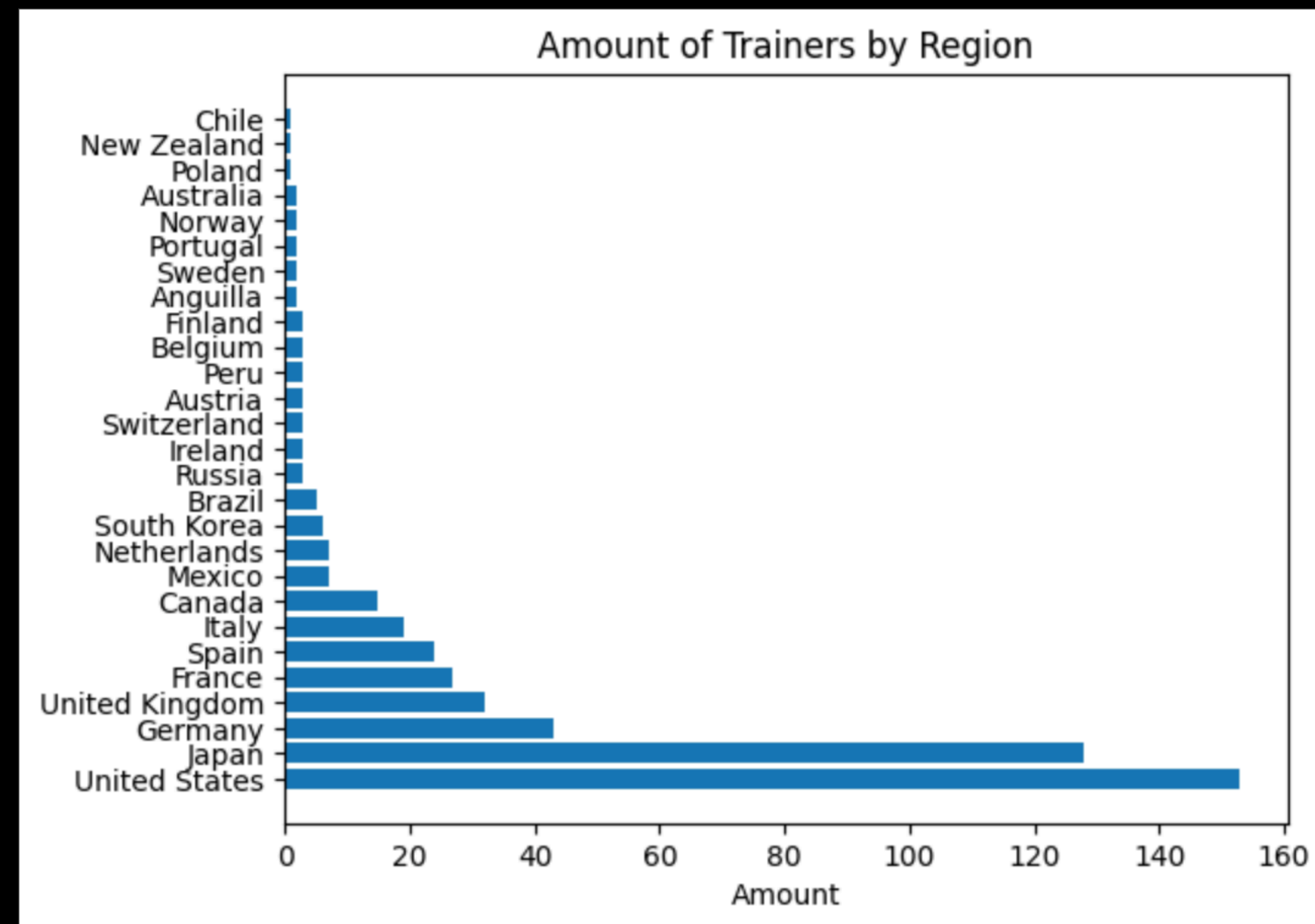


This graph best represents the amount of Pokemon between the levels of 1-10

# Hypothesis

## Understanding the context

Is there a relationship between Pokemon genders and levels? Although it sounds ridiculous, you never know until you visualize it.

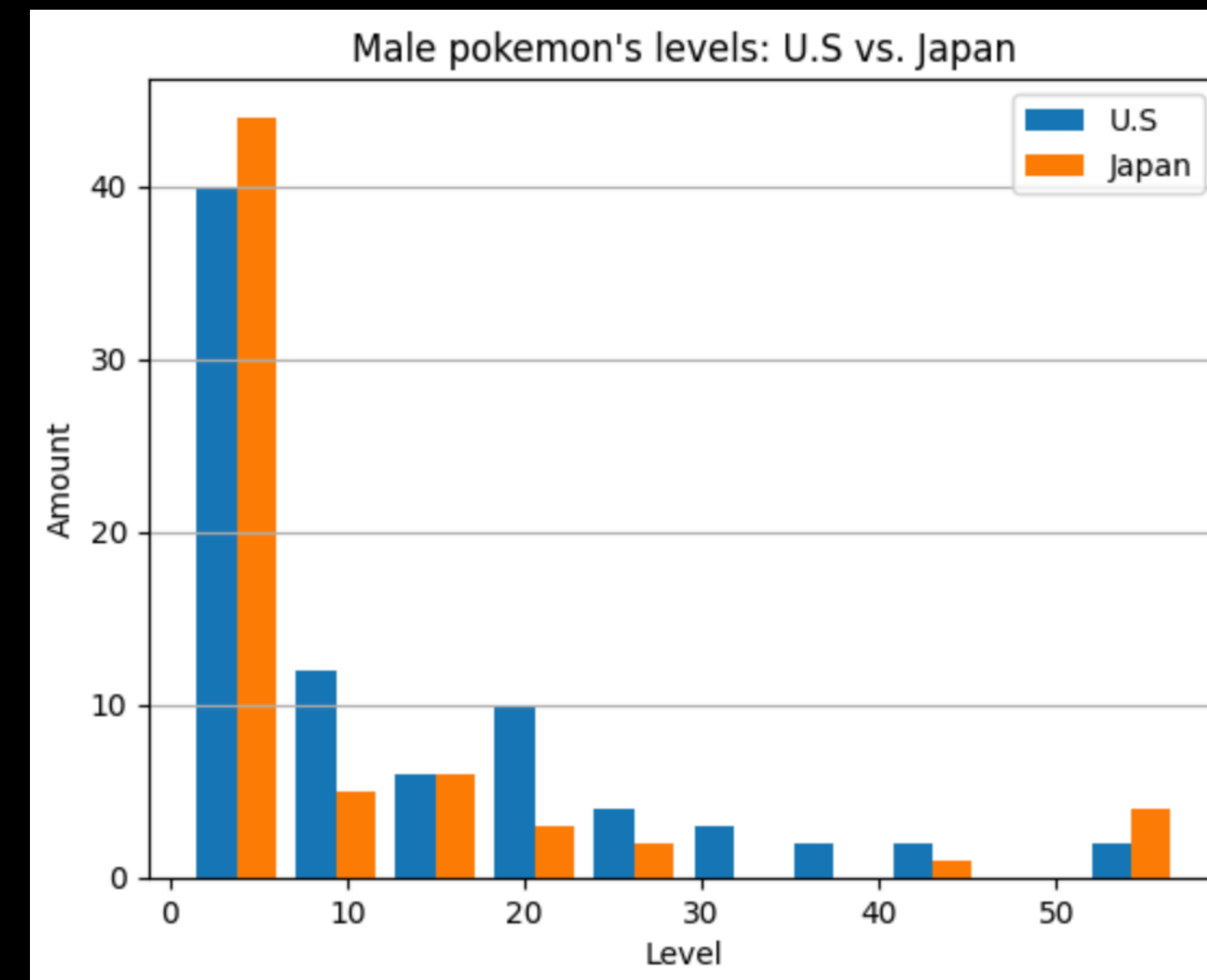
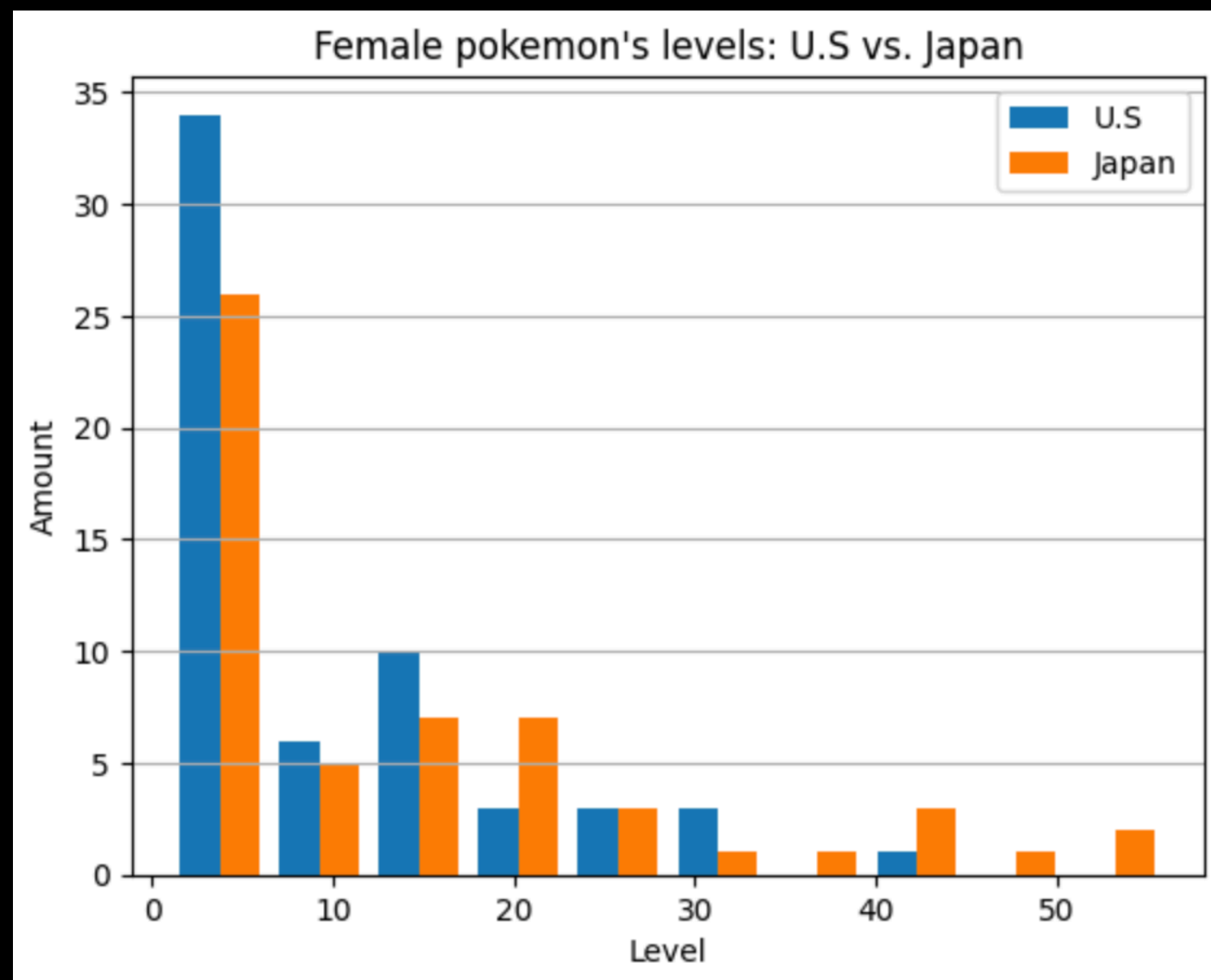


Chose to base comparison  
off of those from Japan and  
U.S

# Visualizing

## Understanding the context

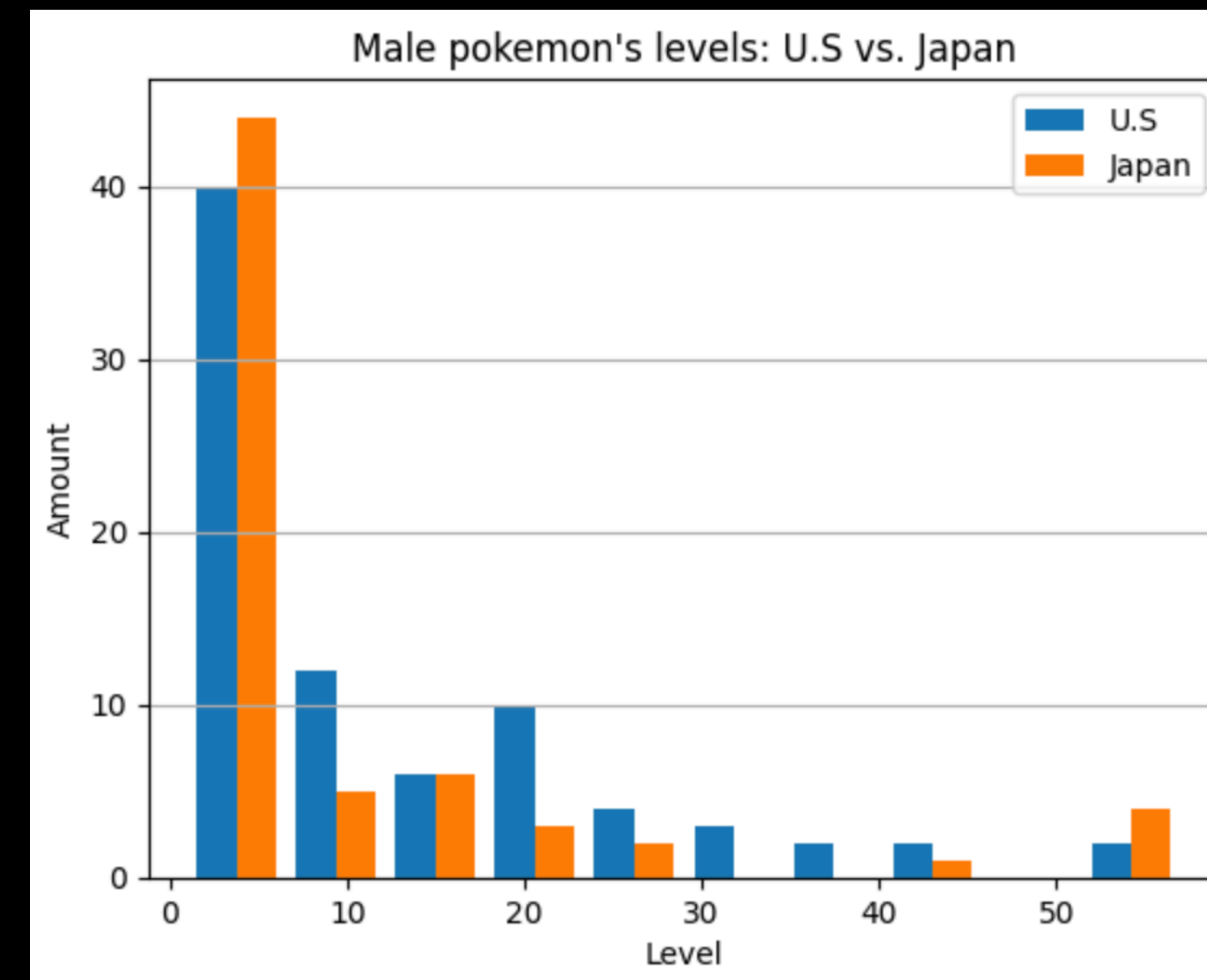
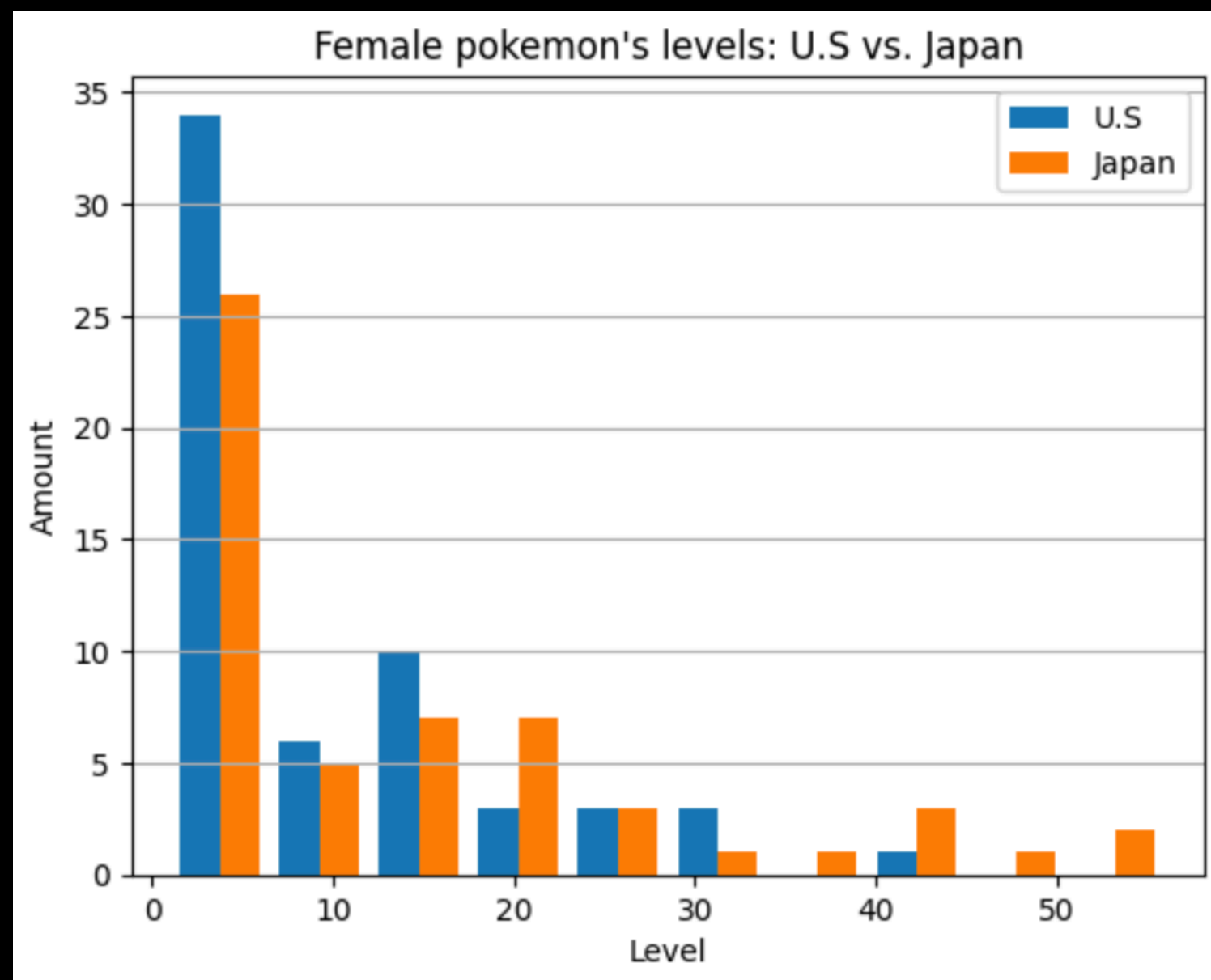
These two graphs represent the male and female Pokemon levels between U.S and Japan.



# Conclusion

What did we get out of this?

We can assume now that (from those two datasets) there is no real trend/relationship of the Pokemon's genders and their levels between the U.S and Japan

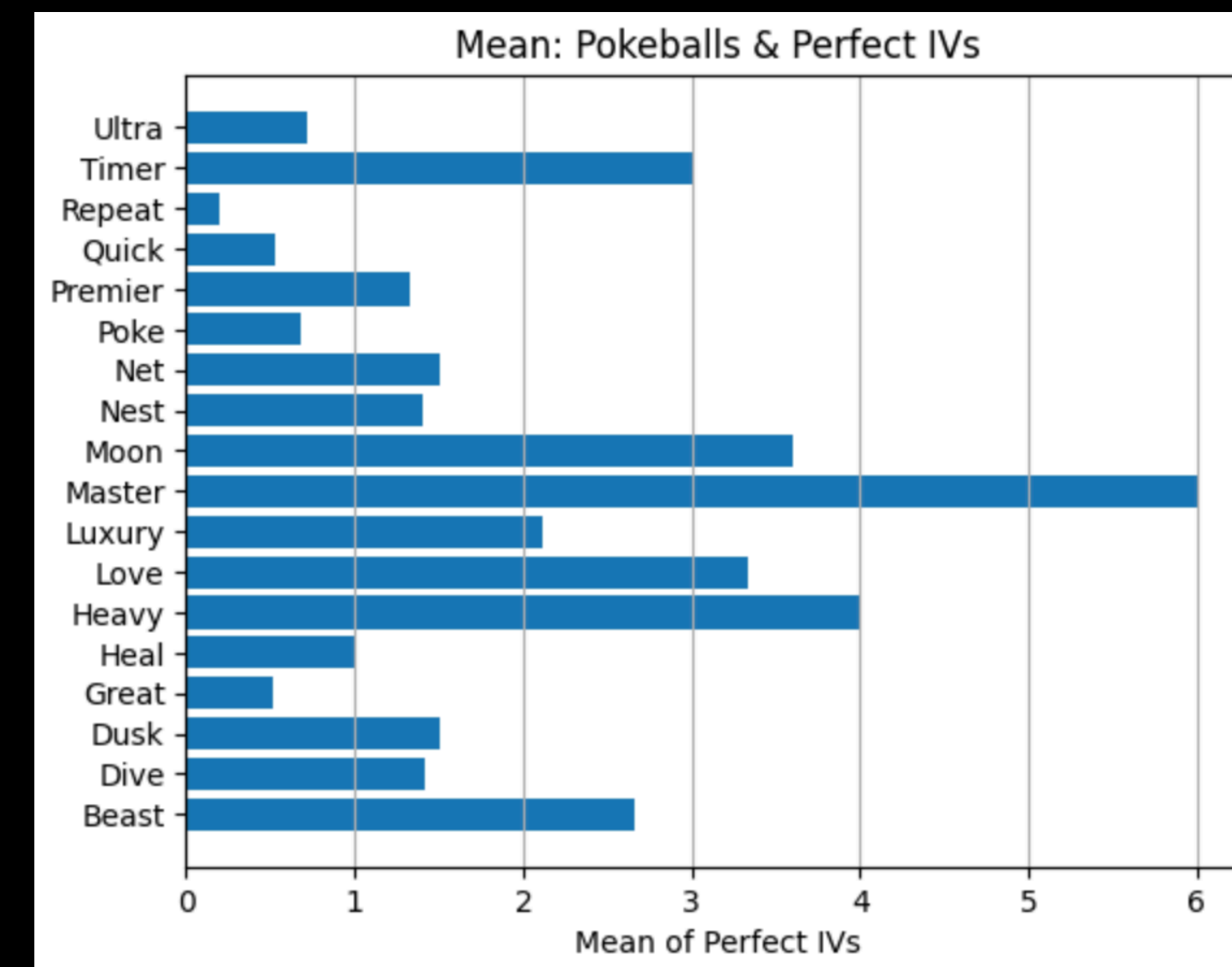
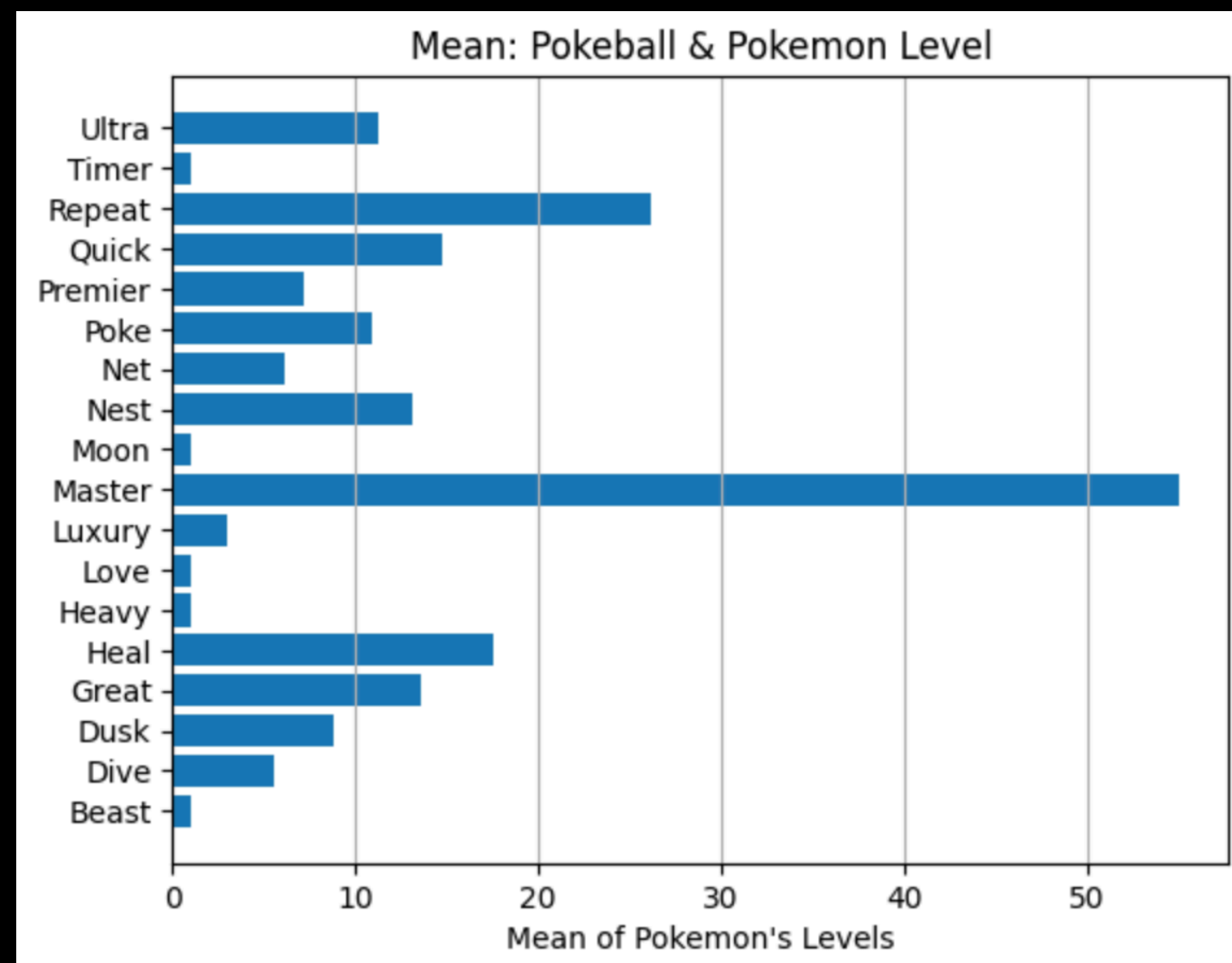




# Hypothesis

## Comparing averages

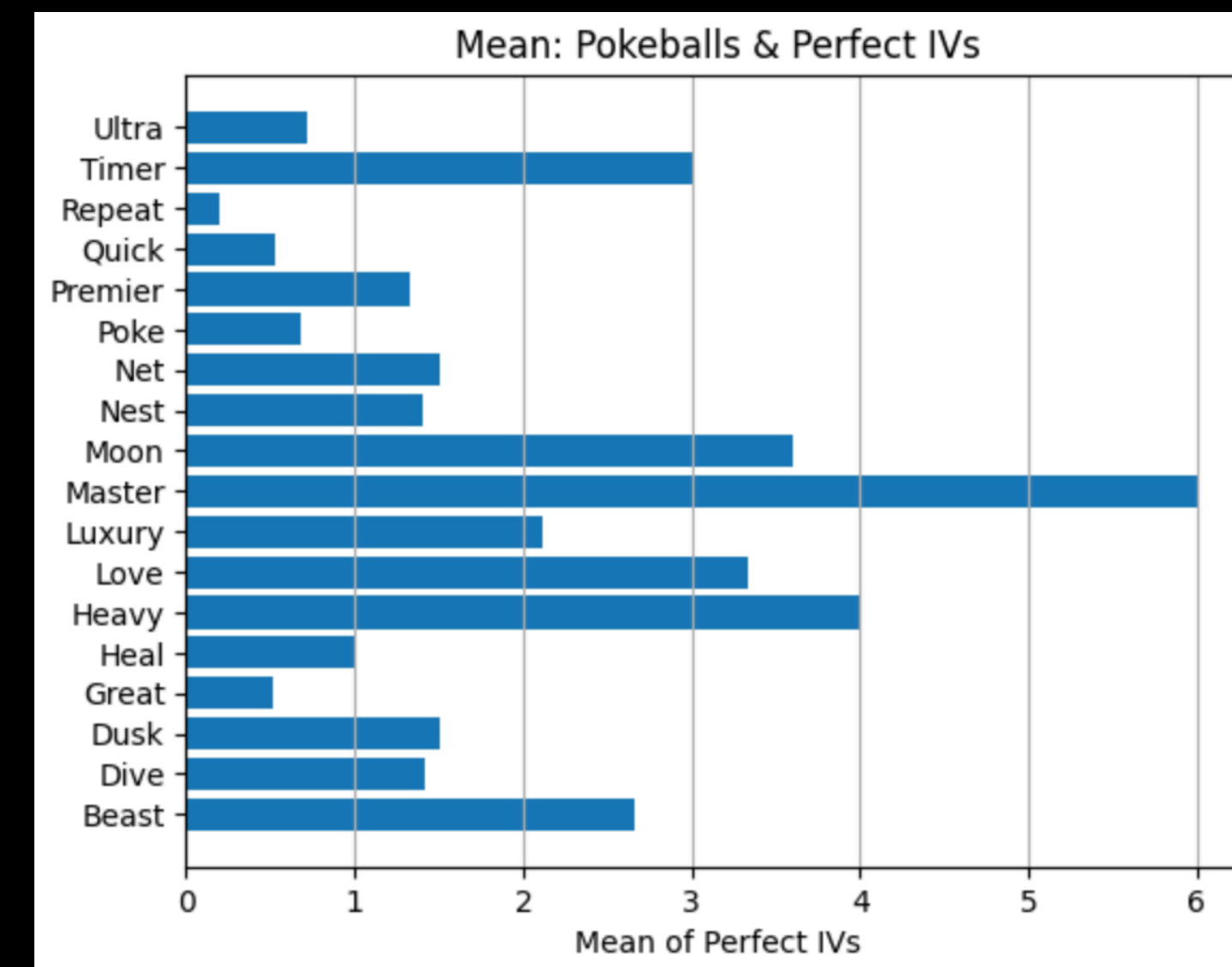
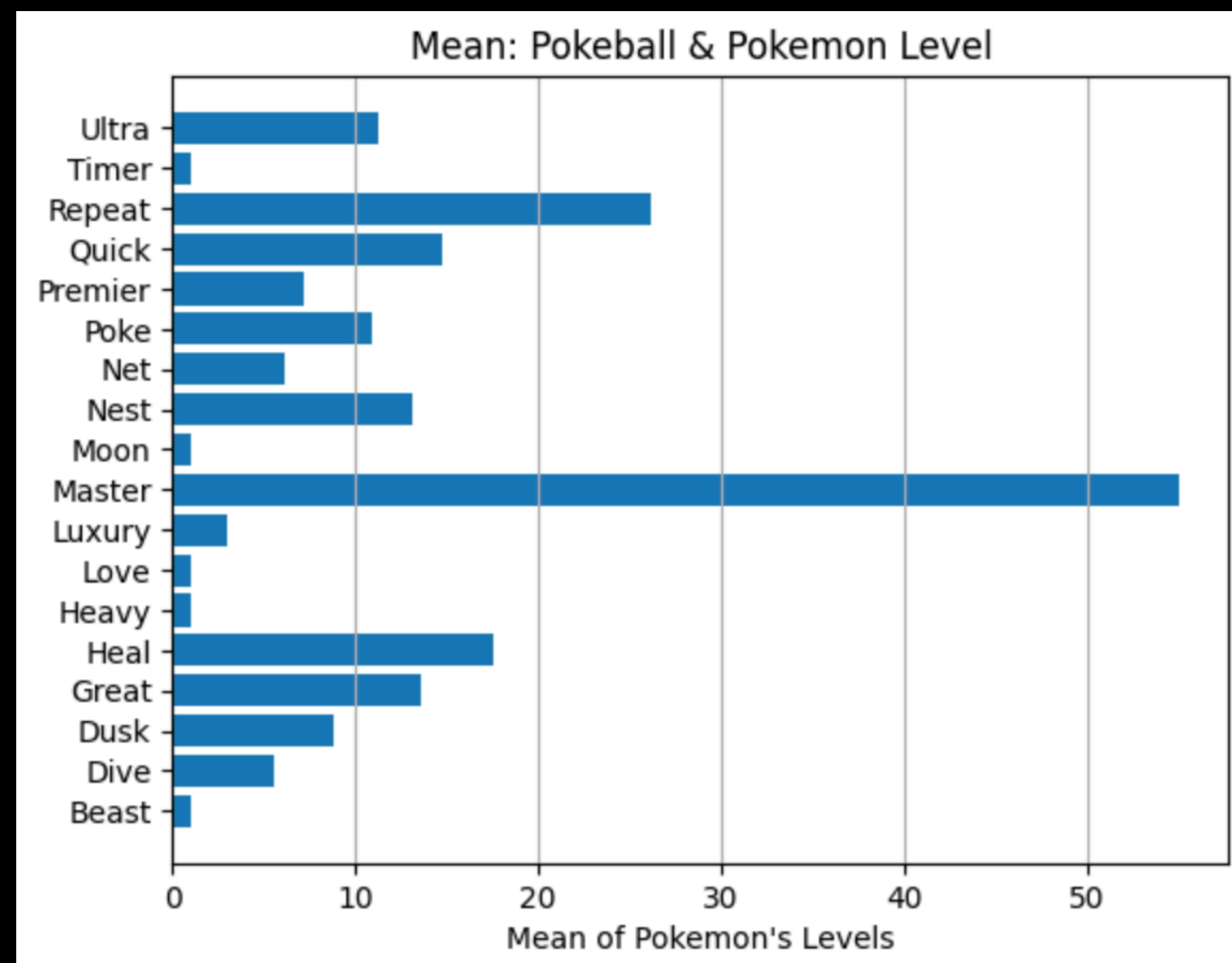
- Can we find out whether or not there's a “best” poke-ball option for catching a Pokemon?
- To test this with the given elements from the dataset, I decided to compare the average (mean) levels and average perfect IVs of each poke-ball.



# Observations

## What do we immediately notice?

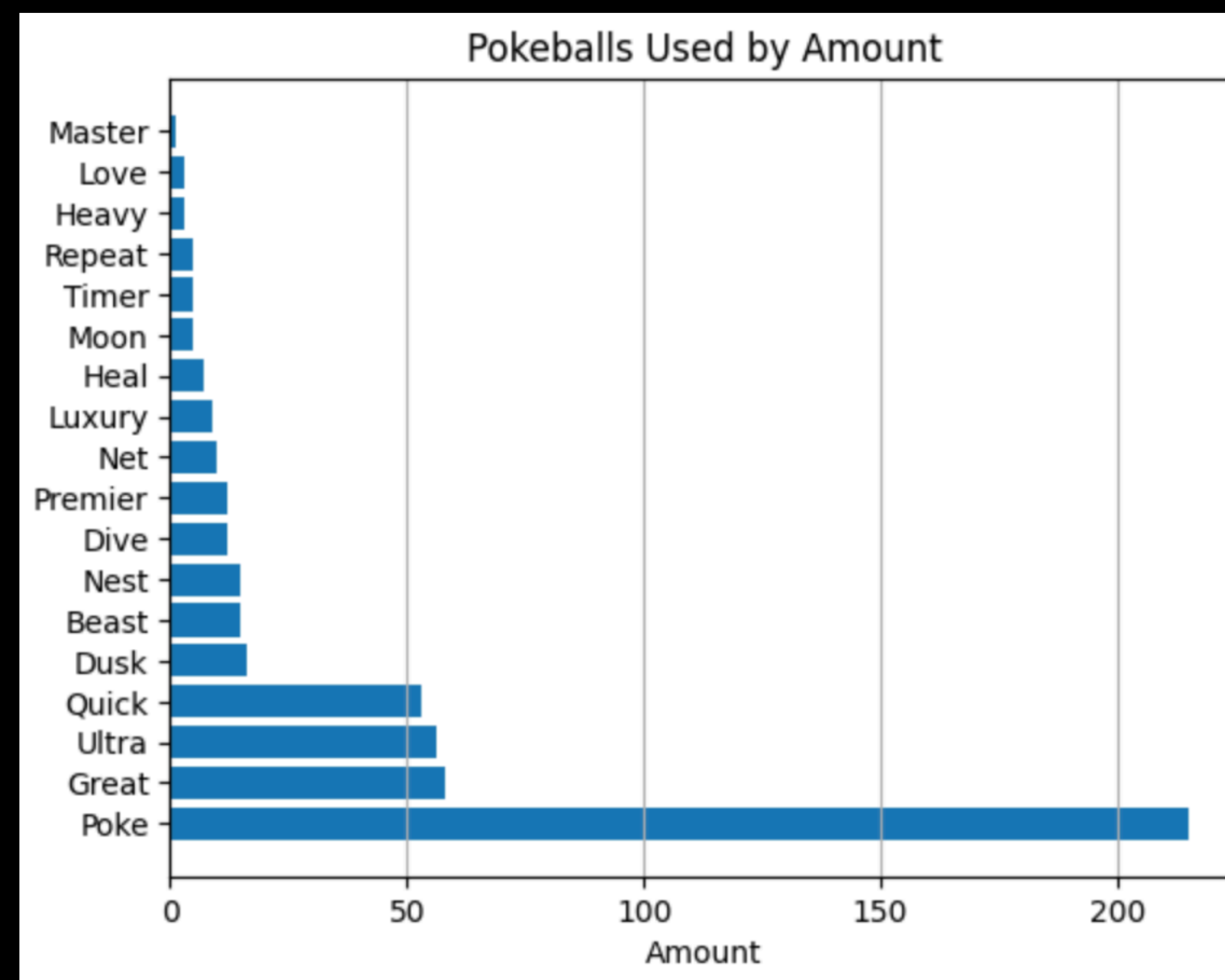
- From these graphs, we immediately notice one clear winner, the Master ball.
- With this, we can assume that the master ball is the best option for catching Pokemon.



# Analyzing the Data

## Visualizing properly

Making assumptions without visualizing all of the necessary data involved is bad practice. Therefore, I visualized the full context for you.



# Conclusion

## Key insights

- Types of graph used to visualize data
- Context of data
- Type of dataset