**Introduction**

I have been a comic book fan for many years and, when I started writing web scrapers for practice, it was only natural that I did one inspired by my passion for Marvel.

Marvel is in the process of creating a sweeping fictional universe created out of a multi-decade set of franchises of movies and buttressed with TV series. This fictional universe is called the Marvel Cinematic Universe. By having a coherent universe, you can develop depth in characters and context for decisions across multiple vantage points.

This also means that you don’t need to spend a lot of time doing backstories or origin stories. They are now introducing characters in other films before their stand-alone franchises, which means that there is a quicker entry into the main movie.

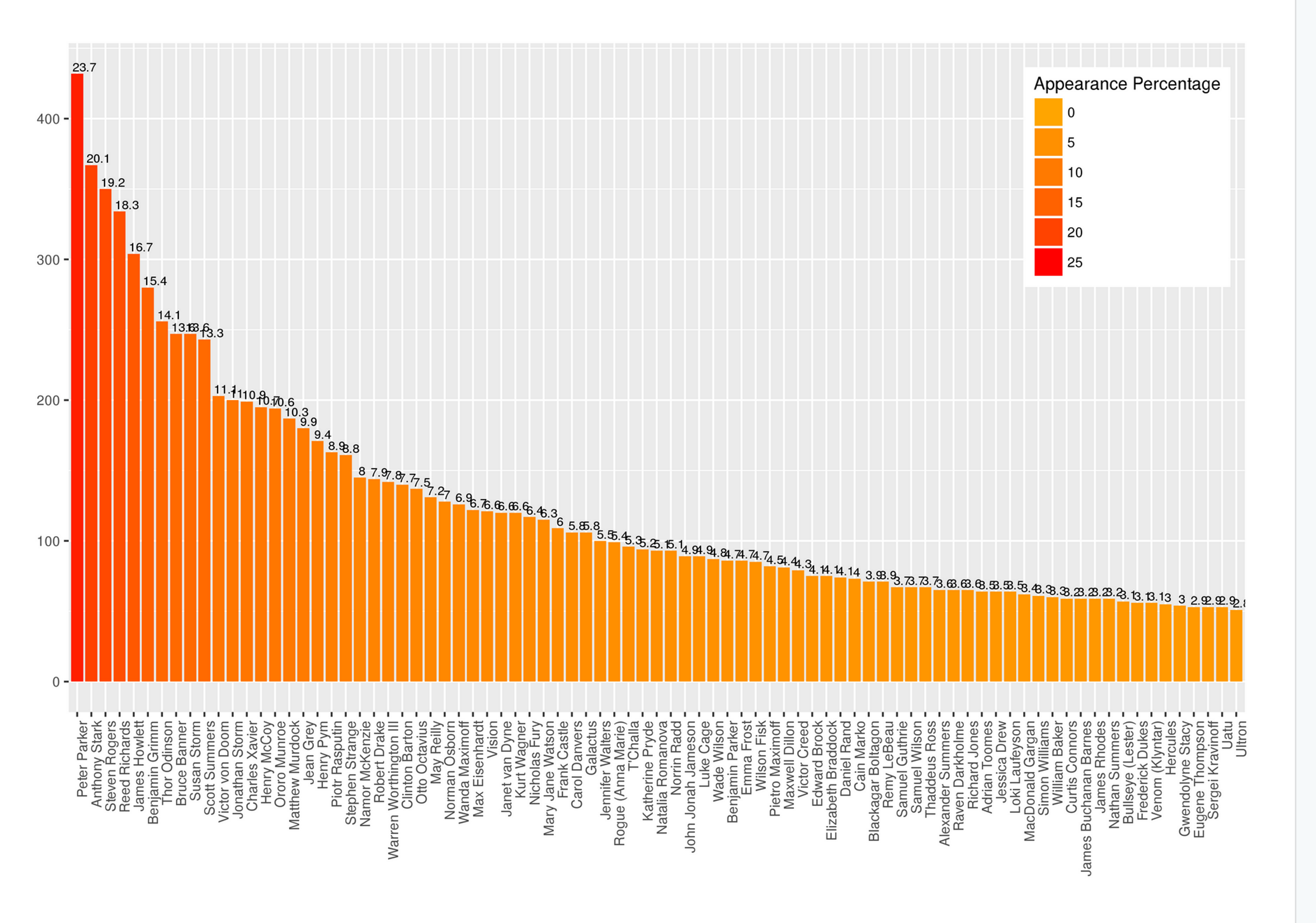
One additional cool point is that the movies are occurring in real-time — meaning that if it has been one year since the last movie in the Marvel Cinematic Universe, one year will have elapsed in the script. This strategy makes the movies seem fresh.

Enough with introduction let’s get down to the business, I can talk for years about marvels. FYI, my favorite super hero is The Hulk in Marvels.

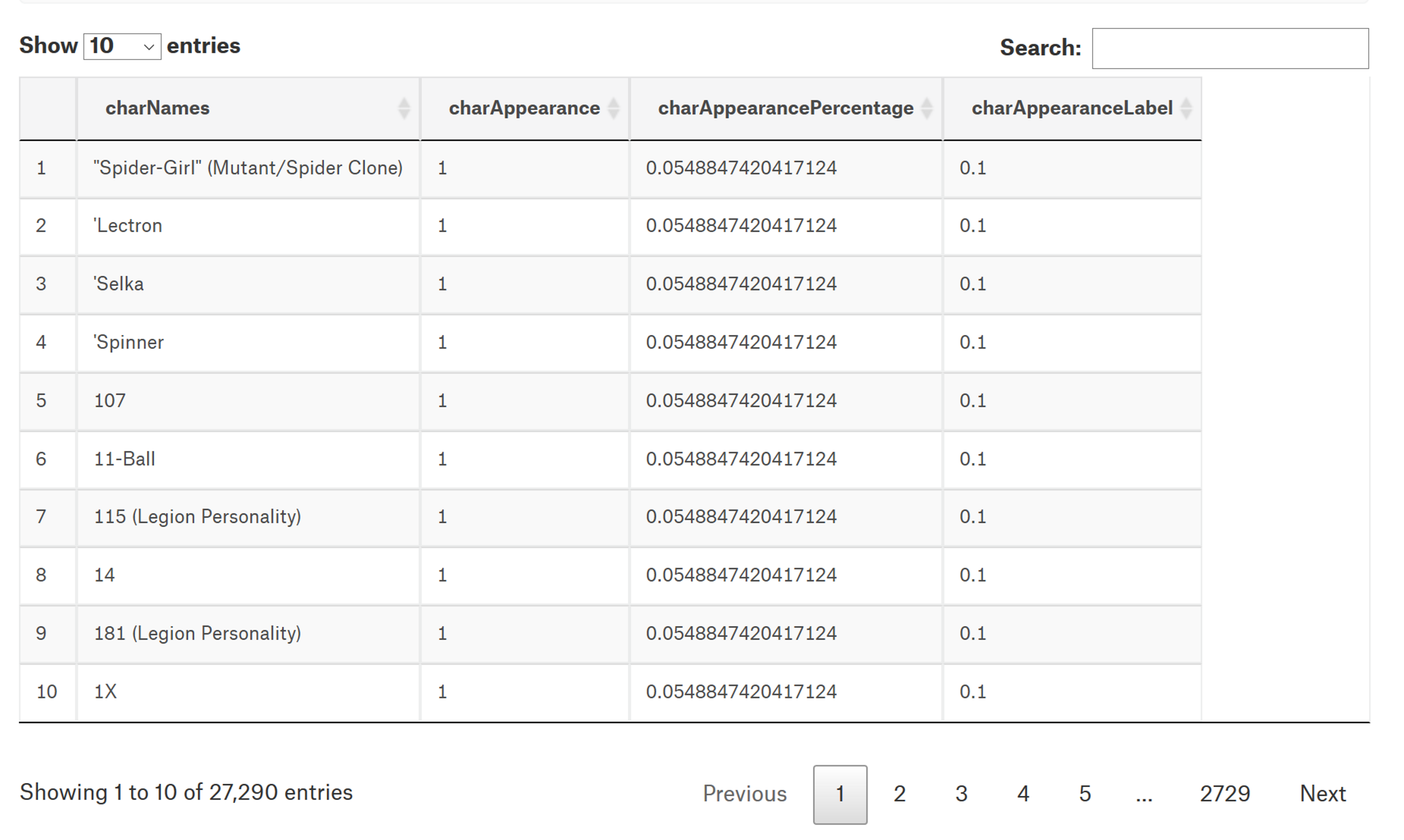
**Dataset**

This dataset has 27.290 rows, each one representing a distinct Marvel character, such as Peter Parker, Tony Stark or Jean Grey. As for columns, there are 1822 of them, one for each Marvel universe. All the cells contain a Boolean value: true if there is a version of that character from that universe or false otherwise. As we are interested of the `Marvel Universe` in which the character exists, we need to select only the `True` case.

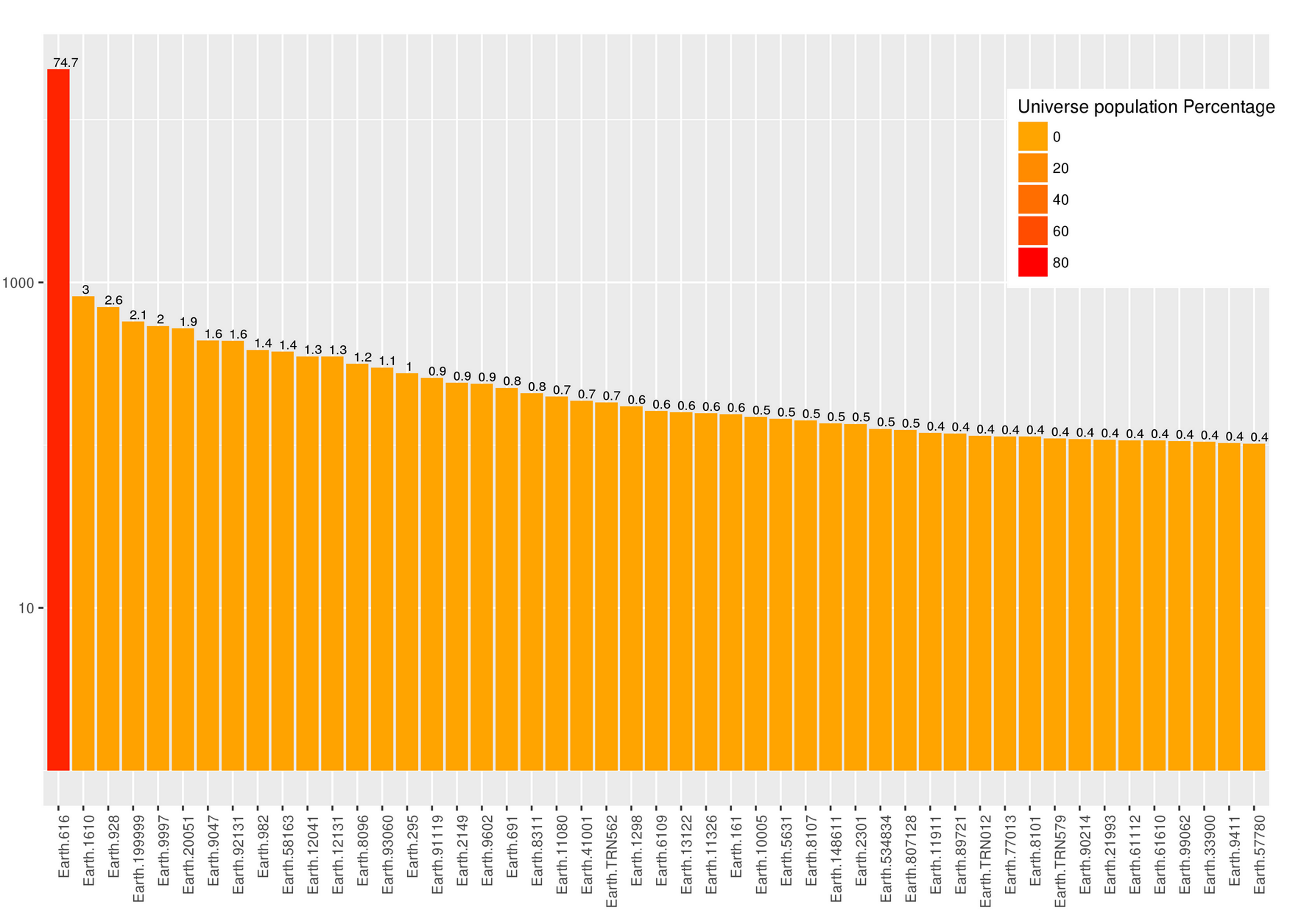
**Now the question is, who is the most frequent Character? I would be disappointed if it’s not Bruse Banner (The Hulk).**



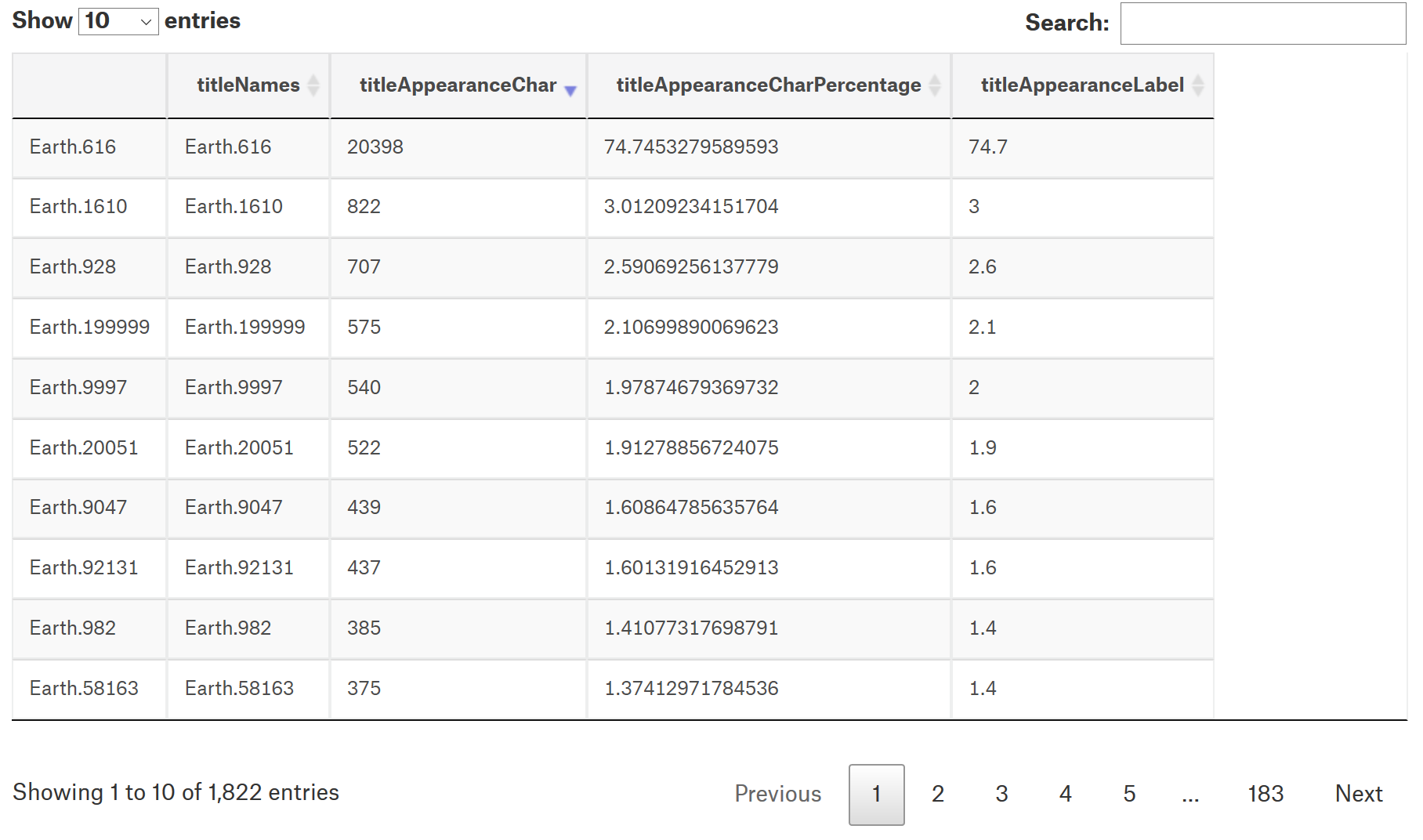
This plot gives the distribution (limited to a number of appearance >50 for visualization purposes) of all characters. For example, Peter Parker appears in more than 400 Universes, which corresponds to ~ 24% of all the Universes (from this dataset).



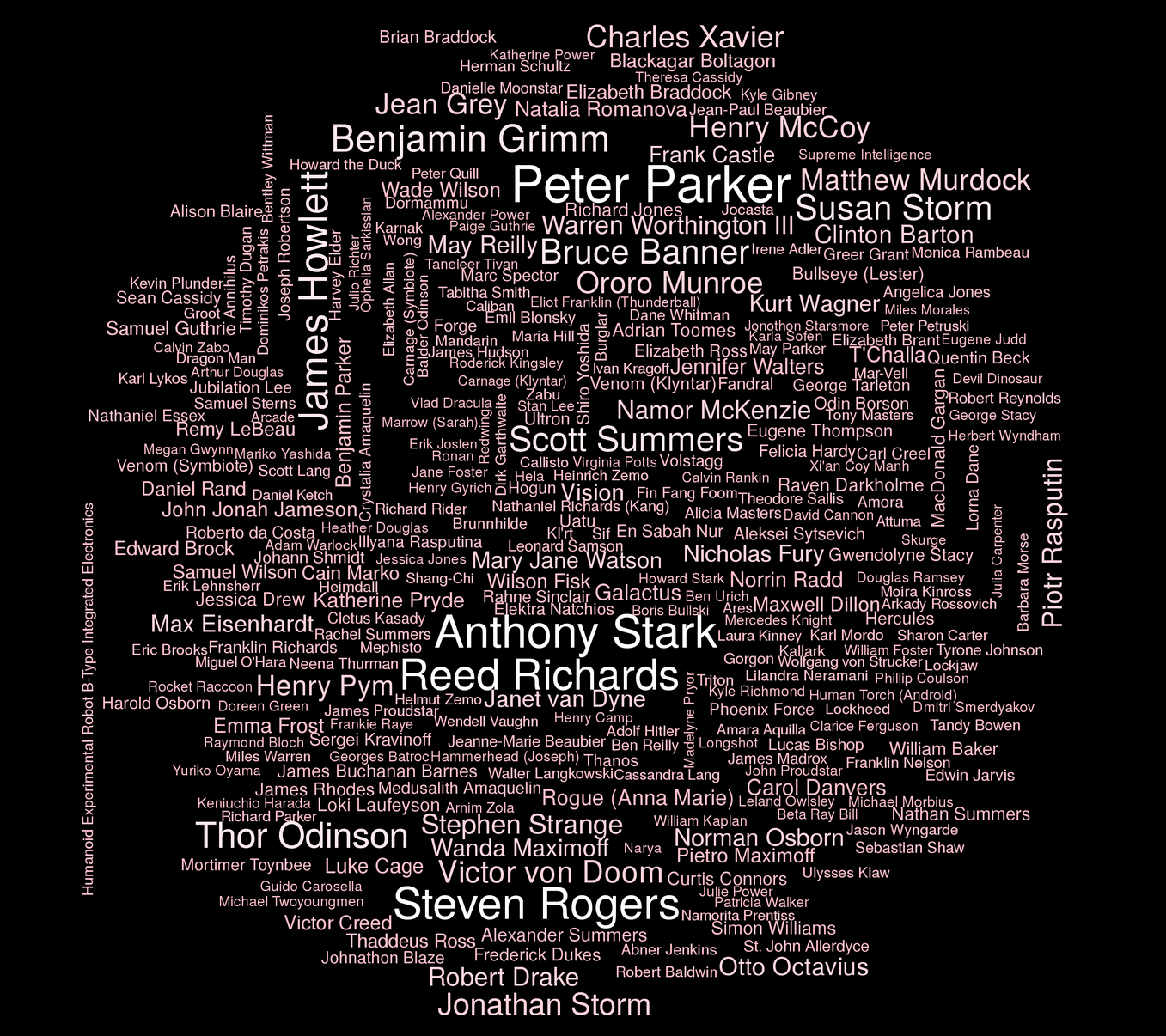
**Which is the most populated Universe?**



This plot gives the distribution of all Universe (limited to a number of Characters existing in it greater than 100 for visualization purposes). Earth.616 is the main Universe with more than 20000 Characters (~75%) (notice the log scale)



**Most Frequent Character**

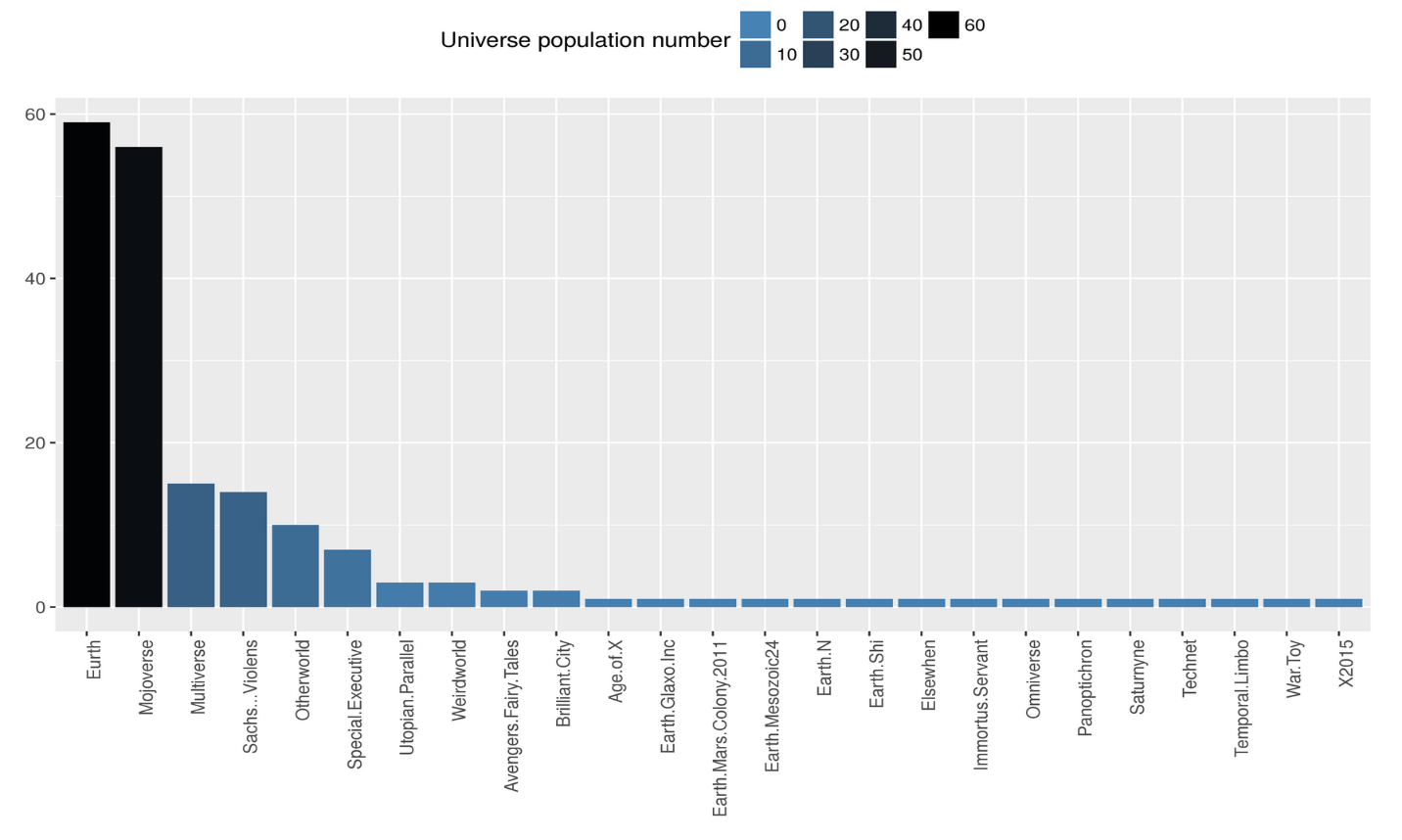
****

**Most Populated Universe**

****

We’ve seen that Earth.616 is the most populated Universe and therefore makes a very skewed distribution. In this section I will just disentangle all the Universe from the Earth-like and look at some distribution.

**Most Populated Universe but not from Earth.XXX**

****