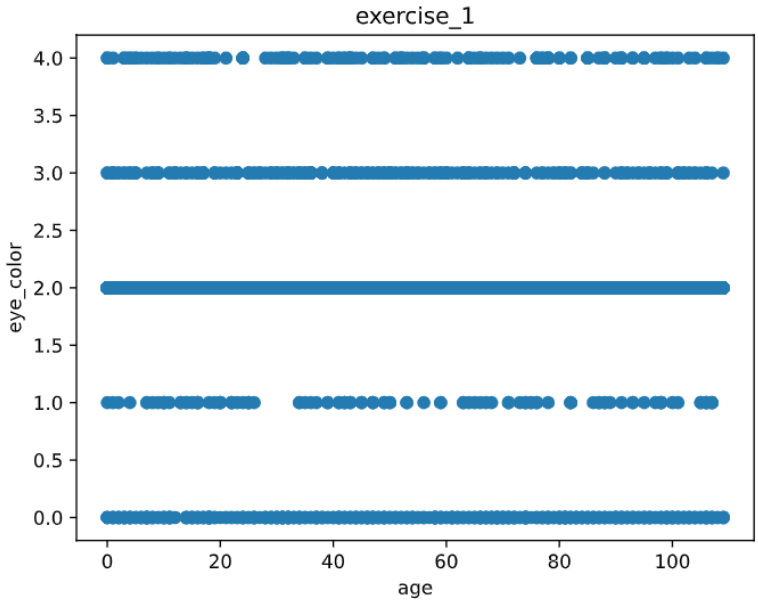
Part 1: data distribution and the law of large numbers

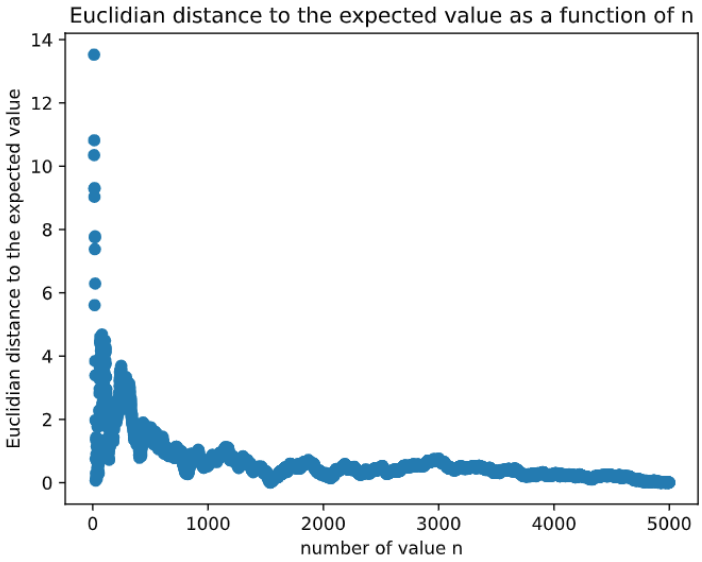
I spent a lot of time trying to figure out what was exactly the matter of the exercise. I wanted first to take as parameters the age of a population and their mass. But I quickly got lost in all random numbers, so I decided change mass parameter to eyes color following real percentage with far less different possible values.

It got better after this change, and I finally understood that goal was to demonstrate law of large numbers by checking difference between expected value and empirical mean and how it converges when size of data is large.

I just got issue on converging around 1 instead of 0 cause I created my array of eyes color with numpy.random.randint() instead of numpy.random.choice() so my custom probabilities for each value wasn't take into account.

After this fix everything went well.




Euclidian distance to the expected value as a function of n
Euclidian distance to the expected value
number of value n
0
2
4
6
8
10
12
14
0
1000
2000
3000
4000
5000