Age Difference in Color Naming

Jiayang Ying

Zixin Huang

Abstract:

Many researches of the relation of age and color naming has done on children and elderlies, many of them agree on there is an age difference in color naming, but the question of whether or not there could be a significant correlation in between or what kind of trend of color naming ability would be as people age, still remains unknown. We are going to research on the question of “Are people able to name more colors as they age?” Moreover, it is noticeable that there might a common theme across all cultures and language, that as people grow older and experience more colors in daily life, they can name more colors. Therefore, it is reasonable to hypothesize that regardless what language people, people able to name more colors as they age. We first calculate the correlation score between age and distinct colors terms across all language groups. We find there is an inconspicuous positive correlation. Then we use linear regression to plot a trend between age and color terms for each language group, in which we find that regardless the language group, generally a positive trend was observed in all population. Finally, we used box plot and bar chart to compare and contrast the variance and mean of average number of color names for different age groups. We find that middle-aged (45 - 65 years old) people actually performance the best in that they use most number of distinct color names. Thus, our hypothesis is only partially right and people, regardless what language they speak, are able to name more colors as they age before they pass middle-age, where after that there is decline in the ability of recognizing and naming different colors.

Introduction:

Color naming is substantially involved in our daily activities; for example, very basic tasks, such as communicating to identify objects, is inseparable from the ability to name color correctly. Since color naming is so crucial, we want to dig into kinds of factors that may have an impact on performance of color naming, where age is one of such factors.

Color naming and color recognition develops very early, according to Bateman, children may begin use color terms around twenty month after they are born.They are able to recognize more and more color as they grow a little older, some 4 years old children are able to use 6 to 18 different color names in English speaking culture. (Bateman, 1915) Some other study, from Joseph and Christina's research on aging and color naming, has shown that older people, over 80 years old, with aging eyes with 41 times less light transmission than normal, younger eyes, which may results in less sensitivity in certains colors. (Hardy, 2005) However, this decline might only results in minor setbacks in distinguish colors and older people's ability of naming and recognizing colors are still relatively well preserved. (Wijk, 1999) Intuitively speaking, age gives individual opportunities for constant practice and repeated exposure to same kind of stimulus usually, then makes a cognitive process more mature; for example, the ability of language speaking is gradually formed as kids practice talking and functions better and better as they grow older. By such reasoning, color naming, as a cognitive process, will also be more fine-tuned and have better performance as age grows. However, any cognitive process is complex such that many factors can influence the process to an unexpected way. People may be able to name more colors when they reach teenager period possibly because during exploration period, brain receives a tremendous new information and is stimulated in a way to optimize color naming skills, but adults color naming-related neurons may already get used to such task, then not be activated in such way and their performance gets worse. We acknowledge possible factors that may lead results to different directions, but for now, our hypothesis is that people can name more colors as they age and we will use data analysis to ascertain if such factors exist.

To further dig into such questions, we worked with World Color Survey dataset (Cook, Kay & Regier, 2005) which was set to inquire into if universal constraints on cross-language color naming exist. This dataset recorded how subjects name 330 common colors using native language across 110 languages. Since this dataset also includes age of each subject and we can easily calculate the amount of distinct colors categories each subject named, we can utilize this dataset to investigate our hypothesis.

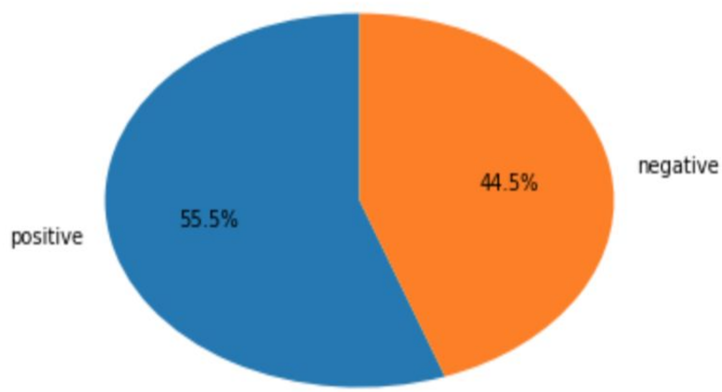
Methods:

We first calculate the correlation score of age and number of distinct colors for every language group. Then we went on to plot for all 110 language groups. Each graph represents the relationship between participants' age vs number of color categories and also applied linear regression on each of them. Then extract all the slopes of the fitted line into this histogram. What we then did is instead compare language groups, we compare different ages. We plotted the avg number of categories vs each age across all languages. We compare 2 data sets, one with eliminate data points with age that has less 15 participants, and the other one is with complete data set. The reason is that, some ages have only 1 or 2 participants, so that would severely skew the plot. We want to see how much p-value would be affected, so that to determine the significance. We also fitted line, find a general trend in this case. We group them together with their age groups: 0 - 10 year old group, 10 - 20 year old group, etc. We plotted the average categories for each groups using box plot to see the variance, mean. Finally plot this bar chart just for the mean, which shows that there's positive trend until they reach middle-age, 40s, and the trend starts to go the opposite direction after around 70 years old. (See Appendix for Python implementation.)

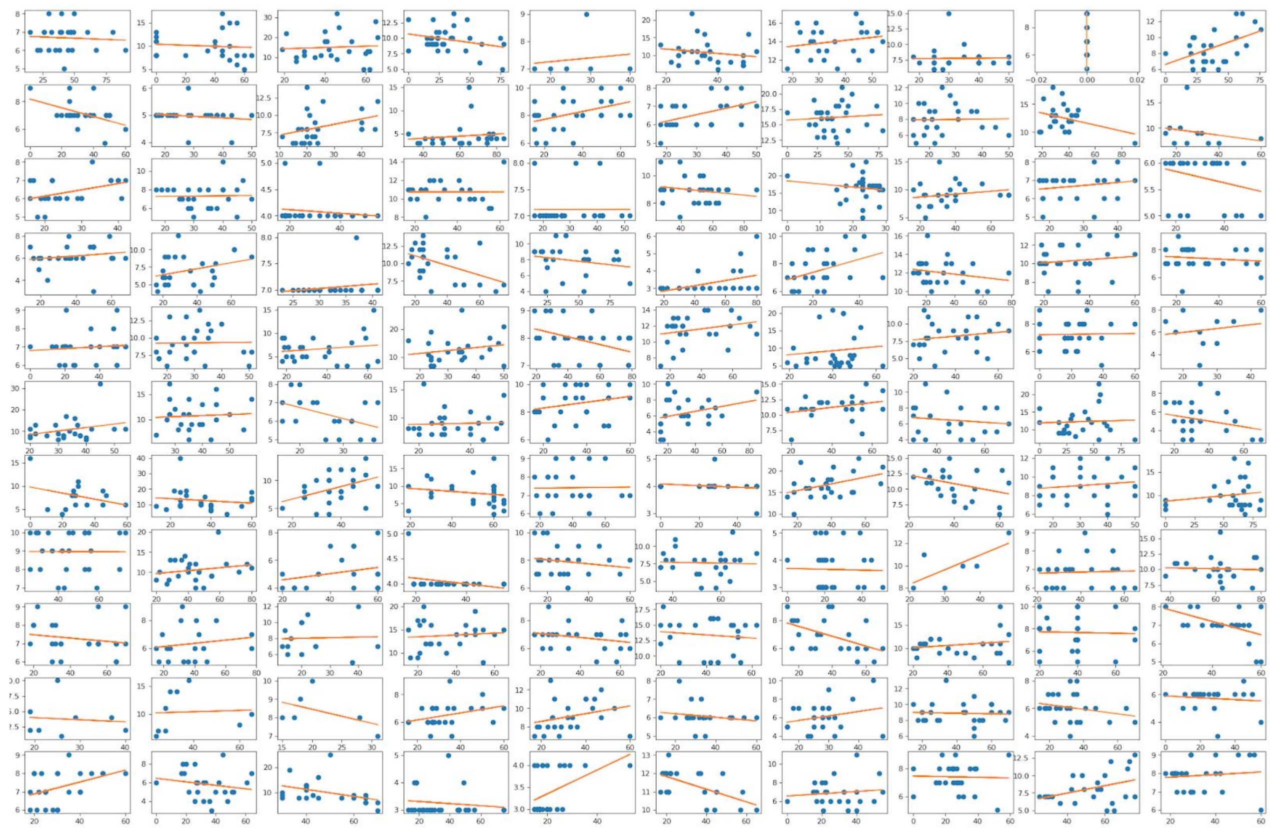
We found there are potential problematic data points in the dataset, such as some language group that age value all equal to 0, this could potentially bias our results. Also, we found, in some language groups, the speaker size is less than 5 people, then it’s not very representative. So we decided to choose 15 participants as a tradeoff choice, if it’s less than 15 participants as not representative, so we have to exclude them in the analysis, if it’s more than 15 participants, it would be more generalizable. In this case, we manage to do a comparison of with and without these exclusions.

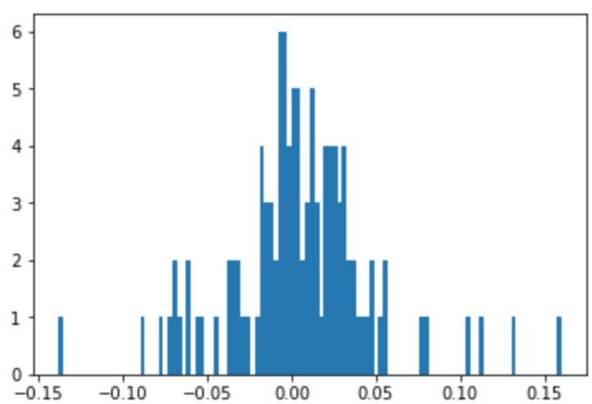
Results:

When we calculate the correlation score of age and number of distinct colors for every language group, we find that 55 to 45 ratio of positive to negative correlation score. This means that across all languages, a positive correlation is observed, but not too obvious.

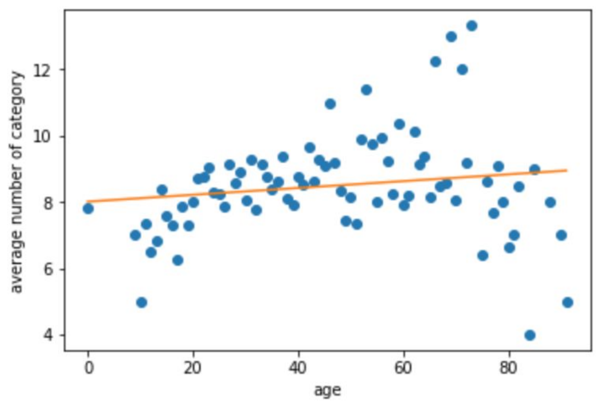


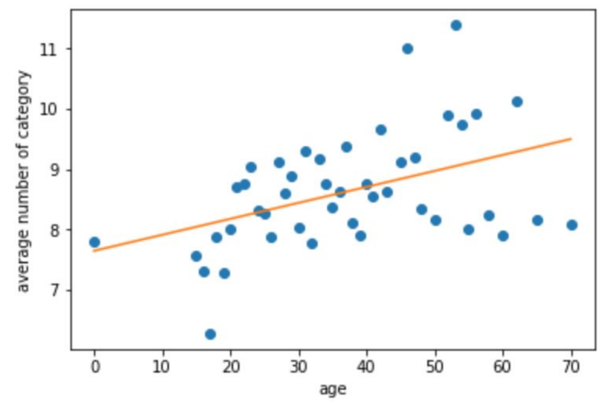
We plot for all 110 language groups. We find majority of the slopes very close to the mean, which suggesting that the differences among all language groups are minimal. So far we’ve found regardless what language they speak, there is a positive correlation. But this positive correlation part seems not very informative. We need to dig deeper.



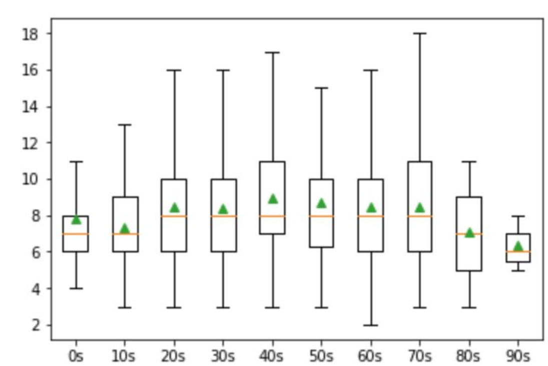


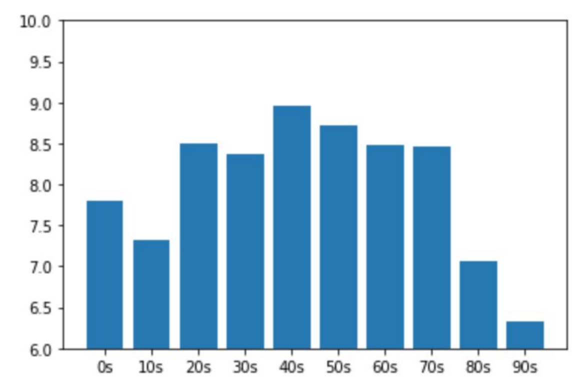
We started to compare different ages instead different language groups. We plotted the average number of categories versus each age across all languages. For example, for all people who are 53 years old they can recognize, in average, 11.4 distinct colors, which is the highest among all ages. We applied linear regression, find a general trend in this case. We find that if we include data points with age that has less 15 participants, we will actually get a p-value of 0.16820, so it would be insignificant. When these data points are excluded since some ages have only 1 or 2 participants, we get a p-value that’s less than 0.05, about 0.026. This means the positive trend is significant. Yet, we think the amount of data points generated in this case here is not enough, because for each age, there’s only one data point.





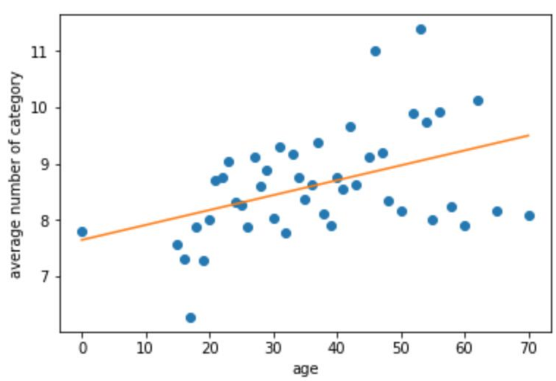
We plotted the average categories for each groups using box plot to see the variance, mean. The variance here may be caused by participant’s language groups, since some language groups have less distinct words for colors and some have way more words in their language.

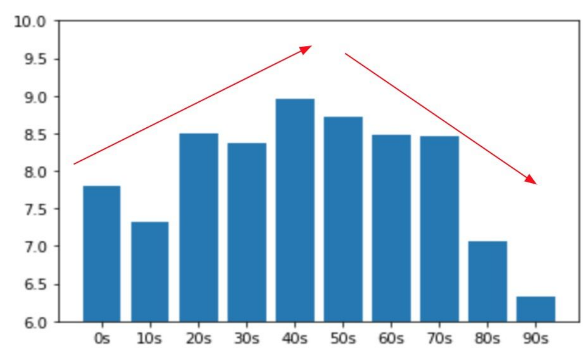




However we notice the mean, these green triangles, may demonstrate a positive in a more obvious way. The mean of each age groups ranging from around 6 - 9 categories. So we zoned in and plot this bar chart, which shows that there’s positive trend until they reach middle-age, 40s, and the trend starts to go the opposite direction.

Why do two graphs, while sharing the same dataset, indicating inconsistent results? The left one is the general positive trend, and the right one is hill-like trend. One potential reason is that in the first analysis, we did not include data from 80s and 90s due to lack of data in 80s and 90s, but in the right graph, we included 80s and 90s data, which bias the general trend to go downwards after middle age. However, if we excluded 80s and 90s data in the right graph, we are still expected to see an upwards trend till middle age and a flat trend after middle age. Hence, such reason is not convincing. A more sensible reason is that, by judging the dot pattern from the right graph, it’s obvious to observe that performance gets diverged after people reaches middle age (40s+), some of them have better performance where most of data centered near 10 categories, some of them have worse performance where most of data centered around 8 categories, when we combine them together, these two effects average each other out and resulting in a neutralized performance, which is around 9 categories. Such result can explain why the value of 60s and 70s in the right graph is less than middle age and therefore, two graphs are essentially compatible.





Divergence after middle age is a very interesting phenomenon. If our hypothesis is correct, we are not expected to observe the divergence, rather, every dots should go upwards. One possible explanation is that it’s true that people can recognize more color when they get older, but old people also tend to be affected by health issues, like optic degeneration, memory loss etc. Such physical problems can directly influence performance on color naming. Elders who keep good health condition can naturally recognize more colors than middle age, but who unfortunately has worse health condition, would perform worse than middle age groups.

Conclusion:

We used data analysis to find the relationship between age and color naming ability and test if our hypothesis is significant. We found there are some flaws of using WCS dataset to test our hypothesis; the result would be more accurate if we can purposely design our experiment to fit our agenda, for example, we can recruit equal size of participants across all age groups across all languages, or we can set up control group for older people who keep good personal health and who do not. Also since the number of color categories is a broad performance indicator, we can come up with a more specific and robust performance indicator, such as color naming distribution across 330 chips. If we find a general and sustained color pattern within particular age group and find general transforming trend across age groups, then we can conclude the role age plays in color naming more confidently and precisely. A recent study by Gibson et al.(2017) found that people like to use more warm colors words(yellow/reds) than cool colors(blue/greens) by analyzing distribution of frequency of 330 color chips across different languages. We can utilize similar method, treating distribution of frequency of color chips across different ages from a single language group as a performance indicator and find patterns across different ages. Despite potential flaws and space for progress, our current results show that as people age to middle age, people gradually have better color naming performance, which verifies our hypothesis, but after middle age, other factors kick in and performance become varied and individualized. Hence, the relationship between age and color naming skill cannot be easily concluded where lots of factors may affect performance.

Reference:

Cook, Richard S., Paul Kay, and Terry Regier. "The world color survey database." Handbook of categorization in cognitive science. 2005. 223-241.

Gibson, Edward, et al. "*Color naming across languages reflects color use*." Proceedings of the National Academy of Sciences 114.40 (2017): 10785-10790.

Hardy, Joseph L., Christina M. Frederick, Paul Kay, and John S. Werner. "Color Naming, Lens Aging, and Grue." *Psychological Science* 16, no. 4 (2005): 321-27. doi:10.1111/j.0956-7976.2005.01534.x.

W. G. Bateman (1915) The Naming of Colors by Children the Binet Test, The Pedagogical Seminary, 22:4, 469-486, DOI: 10.1080/08919402.1915.10533978

Wijk, H., S. Berg, L. Sivik, and B. Steen. "Color Discrimination, Color Naming and Color Preferences in 80-year Olds." *Aging Clinical and Experimental Research* 11, no. 3 (1999): 176-85. doi:10.1007/bf03399660.