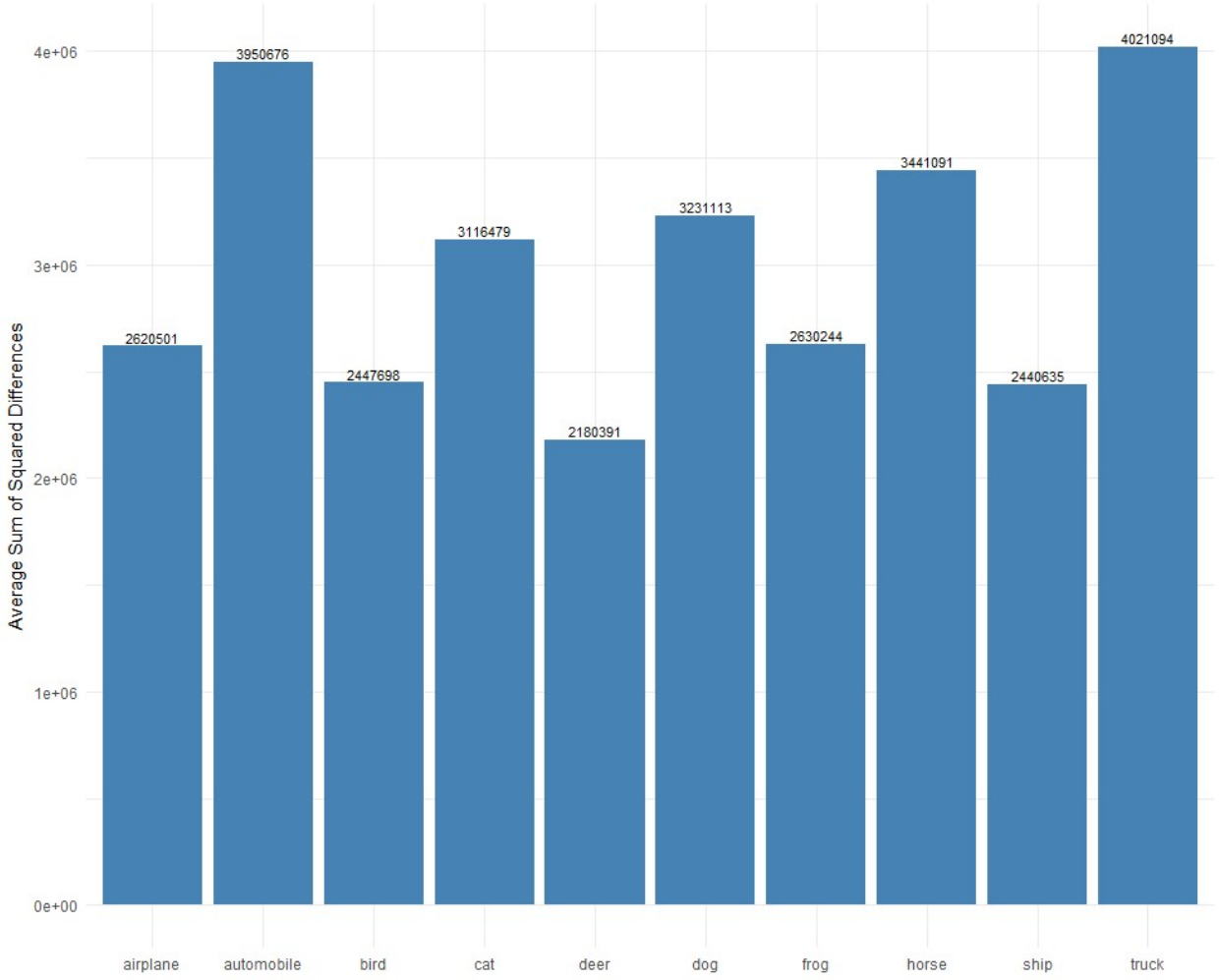
*CIFAR-10 is a dataset of 32x32 images in 10 categories, collected by Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. It is often used to evaluate machine learning algorithms. You can download this dataset from https:// www.cs.toronto.edu/~kriz/cifar.html.*

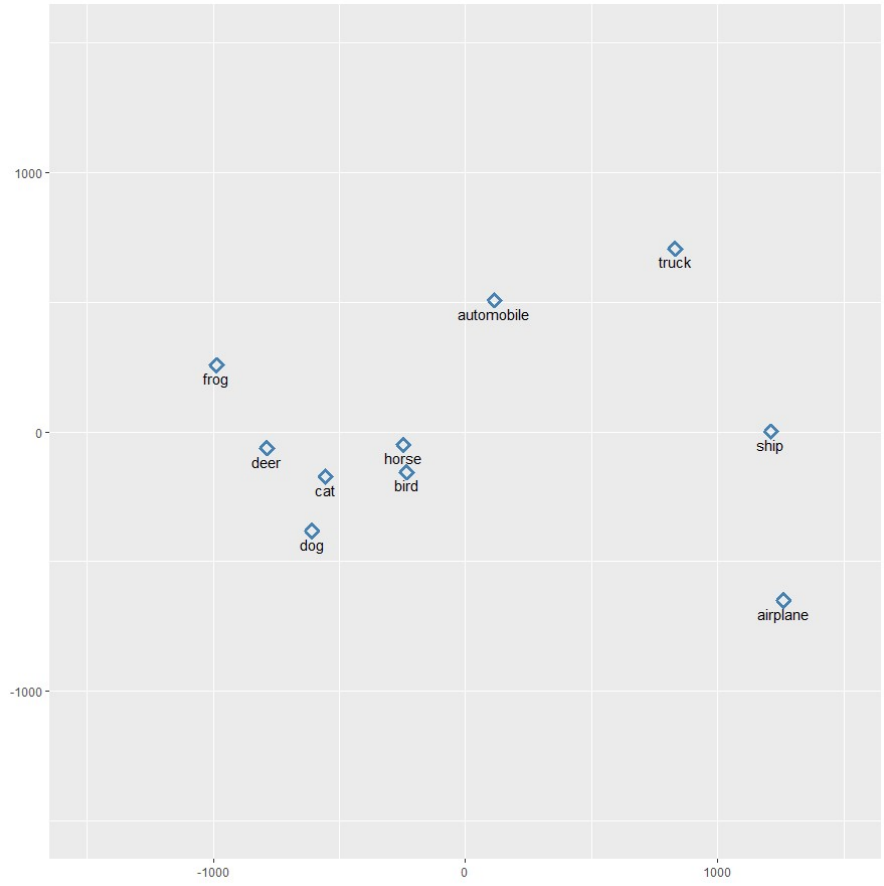
1. *For each category, compute the mean image and the first 20 principal components. Plot the error resulting from representing the images of each category using the first 20 principal components against the category.*

The errors resulting from representing the images of each category using the first 20 principal components are shown below in the bar chart.



*2. Compute the distances between mean images for each pair of classes. Use principal coordinate analysis to make a 2D map of the means of each categories. For this exercise, compute distances by thinking of the images as vectors.*

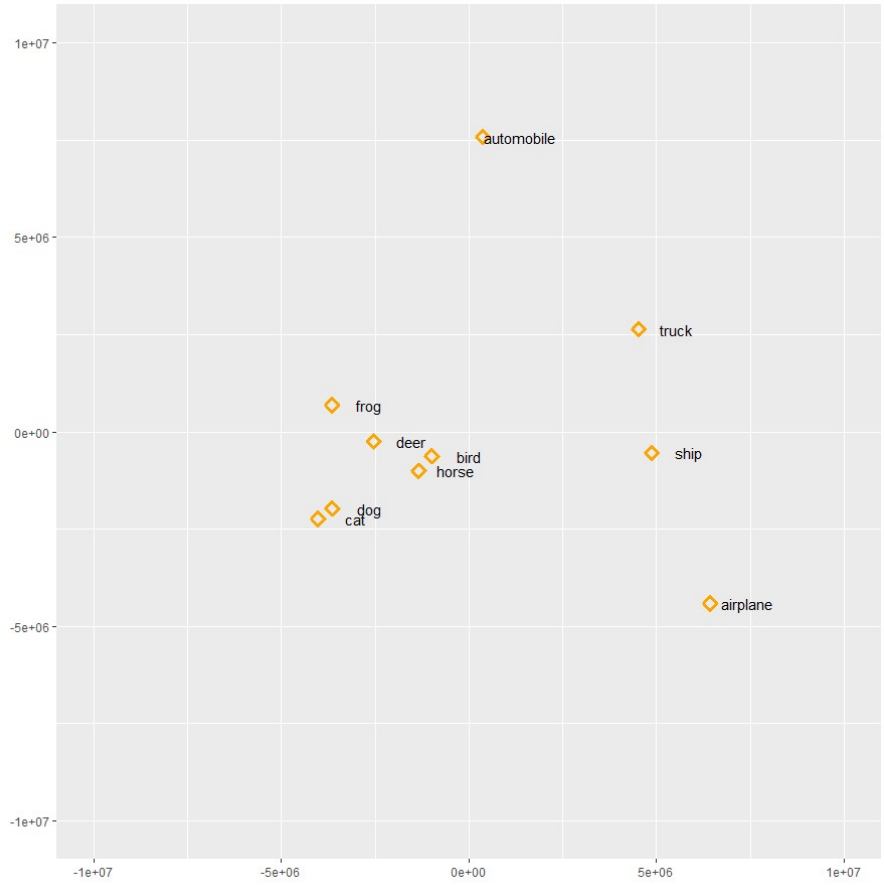
The plot below represents the 2D map of the means of each category. The table below captures the 10 × 10 distance matrix of the classes.




| | airplane | automobile | bird | cat | deer | dog | frog | horse | ship | truck |
|---|---|---|---|---|---|---|---|---|---|---|
| airplane | 0 | 1683.6354 | 1605.0243 | 1905.5353 | 2148.7634 | 1965.2215 | 2445.6797 | 1663.6459 | 945.5411 | 1449.0949 |
| automobile | 1683.6354 | 0 | 886.23675 | 1027.6498 | 1143.0814 | 1216.0794 | 1191.192 | 950.78608 | 1303.4665 | 949.99577 |
| bird | 1605.0243 | 886.23675 | 0 | 517.3115 | 601.25034 | 701.46823 | 913.74752 | 418.27631 | 1557.715 | 1416.6747 |
| cat | 1905.5353 | 1027.6498 | 517.3115 | 0 | 469.79172 | 412.18167 | 677.49197 | 596.37674 | 1851.2145 | 1676.4679 |
| deer | 2148.7634 | 1143.0814 | 601.25034 | 469.79172 | 0 | 617.69714 | 460.51093 | 684.34691 | 2065.6217 | 1830.7409 |
| dog | 1965.2215 | 1216.0794 | 701.46823 | 412.18167 | 617.69714 | 0 | 828.58105 | 843.67209 | 1897.5918 | 1880.2438 |
| frog | 2445.6797 | 1191.192 | 913.74752 | 677.49197 | 460.51093 | 828.58105 | 0 | 948.70402 | 2249.1998 | 1913.2409 |
| horse | 1663.6459 | 950.78608 | 418.27631 | 596.37674 | 684.34691 | 843.67209 | 948.70402 | 0 | 1660.2681 | 1347.3341 |
| ship | 945.5411 | 1303.4665 | 1557.715 | 1851.2145 | 2065.6217 | 1897.5918 | 2249.1998 | 1660.2681 | 0 | 1066.9416 |
| truck | 1449.0949 | 949.99577 | 1416.6747 | 1676.4679 | 1830.7409 | 1880.2438 | 1913.2409 | 1347.3341 | 1066.9416 | 0 |

3. *Here is another measure of the similarity of two classes. For class A and class B, define E(A | B) to be the average error obtained by representing all the images of class A using the mean of class A and the first 20 principal components of class B. Now define the similarity between classes to be (1/2)(E(A | B) + E(B | A)). If A and B are very similar, then this error should be small, because A's principal components should be good at representing B. But if they are very different, then A's principal components should represent B poorly. In turn, the similarity measure should be big. Use principal coordinate analysis to make a 2D map of the classes. Compare this map to the map in the previous exercise? Are they different? Why?*

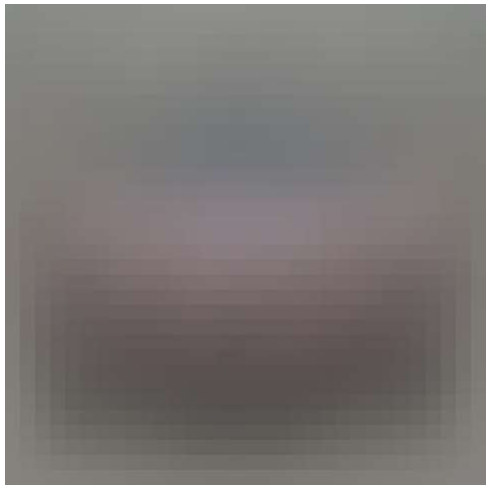
The plot below represents the 2D map of the similarity measures of the categories. The table on the next page captures the 10 × 10 matrix of the similarity measures of the classes. The 2D map for the similarity measures differs from the 2D mean map but there are also interesting similarities. The 2D similarity map has a larger magnitude than the 2D mean map by almost a factor of 1e4. Both maps of the classes in the same general locations. The animal classes (cat, dog, frog, deer, bird and horse) are clustered toward one side and the vehicle classes (automobile, truck, ship and airplane) are spread out to the periphery. In the 2D similarity map, the clustering seems stronger. One possible explanation for the differences is that the similarity measures captures the differences in the principal components among classes as well as the differences in the means. It also suggests that the principal components capture something about being an animal vs. something about being a vehicle that allows for the stronger distinction in the 2D similarity map. This also suggests that using PCA can be very useful in distinguishing images.



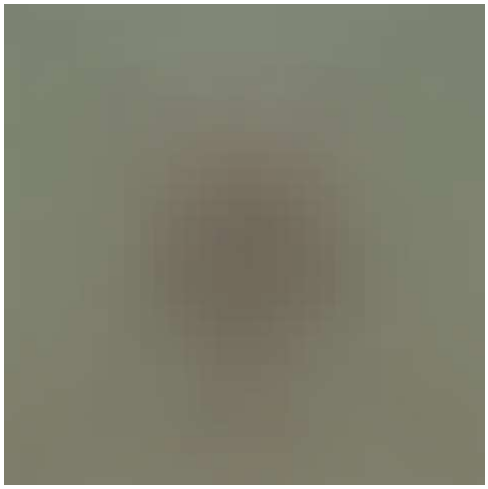
| | airplane | automobile | bird | cat | deer | dog | frog | horse | ship | truck |
|---|---|---|---|---|---|---|---|---|---|---|
| airplane | 2620501 | 13595282 | 12768810 | 14084660 | 13028340 | 13737048 | 13914608 | 12842764 | 11967642 | 12625649 |
| automobile | 14928135 | 3950676 | 13298157 | 14332403 | 12823850 | 14030427 | 13106859 | 13484346 | 13875367 | 13566760 |
| bird | 12595408 | 11797181 | 2447698 | 11615795 | 9950974 | 11265150 | 10479004 | 10822726 | 11854234 | 11775026 |
| cat | 14582854 | 13498352 | 12284536 | 3116479 | 11466418 | 12555929 | 11910715 | 12403594 | 13877270 | 13726415 |
| deer | 12586563 | 11052736 | 9682570 | 10527976 | 2180391 | 10139111 | 9216540 | 9893699 | 11817454 | 11450180 |
| dog | 14348749 | 13312887 | 12046555 | 12671096 | 11195672 | 3231113 | 11645624 | 12174325 | 13654472 | 13672377 |
| frog | 13919154 | 11788717 | 10662662 | 11422092 | 9664474 | 11041451 | 2630244 | 10852800 | 12918403 | 12324767 |
| horse | 13663107 | 12975558 | 11815515 | 12729618 | 11154822 | 12383783 | 11663702 | 3441091 | 13140850 | 12844540 |
| ship | 11786989 | 12363928 | 11849065 | 13203355 | 12080079 | 12861193 | 12729467 | 12137276 | 2440635 | 11404790 |
| truck | 14024669 | 13633722 | 13348172 | 14633747 | 13289868 | 14463473 | 13711889 | 13423791 | 12983210 | 4021094 |

## Appendix: Mean Images of the different classes

airplane



automobile

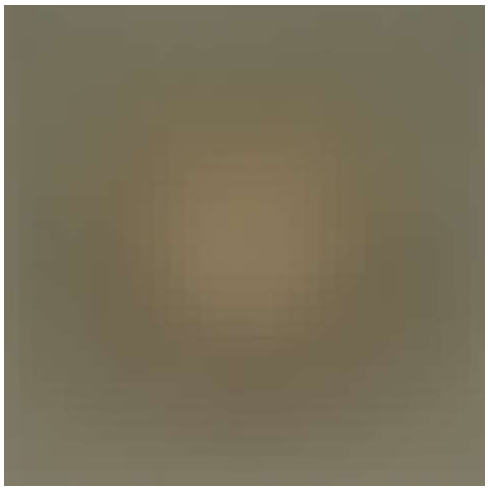


bird

cat

deer

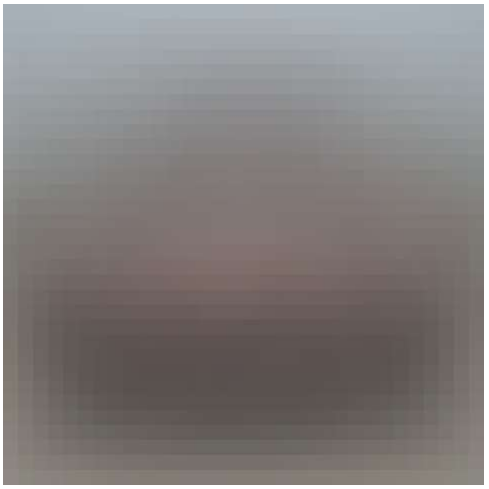
dog



frog

horse

ship



truck