

# **Image Color Quantization and Image Compression**

**CS 6301 – Practical Aspects of Data Science  
Summer 2020**

**Rajarshi Chattopadhyay  
RXC170010**

## Introduction

This project is to perform color quantization of an image using k-means clustering with different values of k and image compression using Principal Component Analysis with different number of principal components.

## Image

An image is downloaded from the internet using the URL.

In this project an image of a Lamborghini car has been used: [https://bringatrailer.com/wp-content/uploads/2019/08/1998\\_lamborghini\\_diablo\\_vt\\_roadster\\_156686808165ef66e7dff9f987625-e1568242425520.jpg?w=400](https://bringatrailer.com/wp-content/uploads/2019/08/1998_lamborghini_diablo_vt_roadster_156686808165ef66e7dff9f987625-e1568242425520.jpg?w=400)

The image is downloaded and saved in a temporary folder created 'files/image.jpg'



The size of the image is noted as 37255 bytes ~ 36.4 KB

Dimension of image = 267 x 400

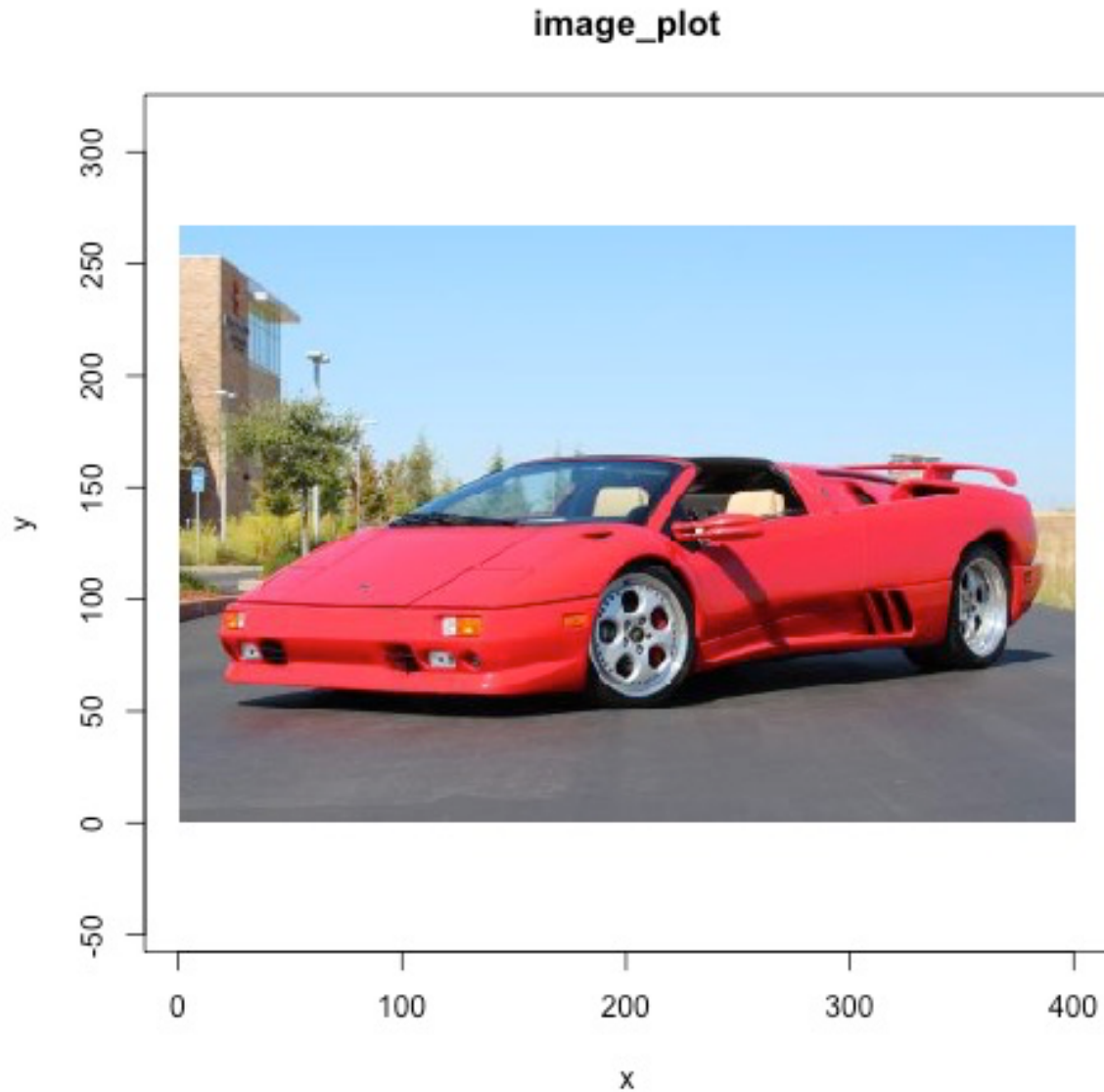
## Package used

The 'jpeg' package has been used for reading .jpg file.

The image is thus represented as 3 267 x 400 matrices as an array with each matrix corresponding to the RGB color value scheme.

## Image Plot

An image data frame is created using its dimension and RGB values.  
The image is plotted and stored at 'files/image\_plot.jpg'



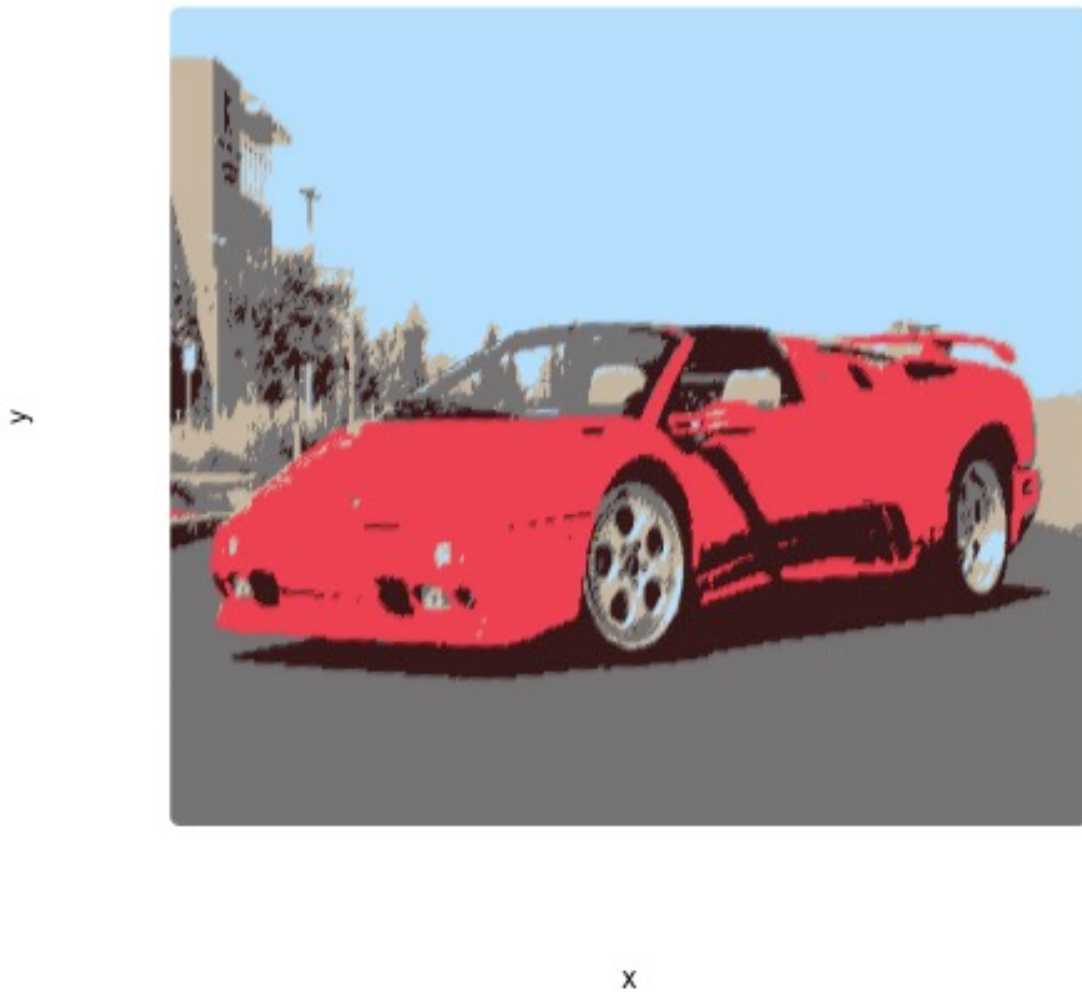
The size of the image plot is 39245 bytes ~ 38.3 KB

## Image Color Quantization using K-means Clustering

The image color clustering is done using k-means clustering for different k values 5,10,15,20,25.

### *Image Plot after clustering with $k=5$*

The image is plotted and stored at 'files/image\_plot\_k5.jpg' of size 30088 bytes ~ 29.4 KB



***Image Plot after clustering with  $k=10$***

The image is plotted and stored at 'files/image\_plot\_k10.jpg' of size 31588 bytes ~ 30.8 KB

y



x

***Image Plot after clustering with  $k=15$***

The image is plotted and stored at 'files/image\_plot\_k15.jpg' of size 31908 bytes ~ 31.2 KB

y



x

***Image Plot after clustering with  $k=20$***

The image is plotted and stored at 'files/image\_plot\_k20.jpg' of size 31872 bytes ~ 31.1 KB

y



x

***Image Plot after clustering with  $k=25$***

The image is plotted and stored at 'files/image\_plot\_k25.jpg' of size 31489 bytes ~ 30.8 KB

y



x



# Image Compression using Principal Component Analysis

Principal Component Analysis of the image is done. The image is then reconstructed with the projections of the data using different number of the principal components:

25,50,100,150,200

A snapshot of the importance of the components is as below:

	PC126	PC127	PC128	PC129	PC130	PC131	PC132	PC133	PC134
Standard deviation	0.02022	0.01952	0.01943	0.01885	0.01809	0.01786	0.0178	0.01773	0.01753
Proportion of Variance	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.0000	0.00000	0.00000
Cumulative Proportion	0.99995	0.99995	0.99995	0.99995	0.99996	0.99996	1.0000	0.99996	0.99996
	PC135	PC136	PC137	PC138	PC139	PC140	PC141	PC142	
Standard deviation	0.01729	0.01711	0.01637	0.01598	0.01574	0.01551	0.01518	0.01471	
Proportion of Variance	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
Cumulative Proportion	0.99996	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	
	PC143	PC144	PC145	PC146	PC147	PC148	PC149	PC150	
Standard deviation	0.01437	0.01418	0.01379	0.01329	0.01312	0.01287	0.01275	0.01249	
Proportion of Variance	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
Cumulative Proportion	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	
	PC151	PC152	PC153	PC154	PC155	PC156	PC157	PC158	PC159
Standard deviation	0.01226	0.01198	0.01168	0.01144	0.01121	0.0111	0.01085	0.01059	0.0102
Proportion of Variance	0.00000	0.00000	0.00000	0.00000	0.00000	0.0000	0.00000	0.00000	0.0000
Cumulative Proportion	0.99998	0.99998	0.99999	0.99999	0.99999	1.0000	0.99999	0.99999	1.0000
	PC160	PC161	PC162	PC163	PC164	PC165	PC166	PC167	
Standard deviation	0.009943	0.009682	0.009378	0.009313	0.009031	0.008834	0.008782	0.00836	
Proportion of Variance	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	
Cumulative Proportion	0.999990	0.999990	0.999990	0.999990	0.999990	0.999990	0.999990	0.999990	0.99999
	PC168	PC169	PC170	PC171	PC172	PC173	PC174		
Standard deviation	0.008309	0.008153	0.008002	0.007566	0.007452	0.007312	0.007164		
Proportion of Variance	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000		
Cumulative Proportion	0.999990	0.999990	0.999990	0.999990	0.999990	0.999990	1.000000		
	PC175	PC176	PC177	PC178	PC179	PC180	PC181	PC182	
Standard deviation	0.007024	0.006728	0.006667	0.006579	0.006393	0.00622	0.006078	0.005769	
Proportion of Variance	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	
Cumulative Proportion	1.000000	1.000000	1.000000	1.000000	1.000000	1.00000	1.000000	1.000000	
	PC183	PC184	PC185	PC186	PC187	PC188	PC189		
Standard deviation	0.005648	0.005587	0.005515	0.005224	0.005077	0.005015	0.004868		
Proportion of Variance	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000		
Cumulative Proportion	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000		
	PC190	PC191	PC192	PC193	PC194	PC195	PC196	PC197	
Standard deviation	0.004648	0.004389	0.00437	0.004356	0.00407	0.003976	0.00383	0.003692	
Proportion of Variance	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.00000	0.000000	
Cumulative Proportion	1.000000	1.000000	1.00000	1.000000	1.00000	1.000000	1.00000	1.000000	
	PC198	PC199	PC200	PC201	PC202	PC203	PC204		
Standard deviation	0.003677	0.003427	0.003248	0.003174	0.003046	0.002917	0.002814		
Proportion of Variance	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000		
Cumulative Proportion	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000		
	PC205	PC206	PC207	PC208	PC209	PC210	PC211		
Standard deviation	0.002774	0.002586	0.002526	0.002458	0.002394	0.002369	0.002173		
Proportion of Variance	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000		
Cumulative Proportion	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000		

This shows that around 174 Principal Components cover most of the information of the image.

### ***Image Plot with 25 PCs***

The image is reconstructed and stored at 'files/image\_plot\_pc25.jpg'



The size of the image plot with 25 PCs is 23081 bytes ~ 22.5 KB

### ***Image Plot with 50 PCs***

The image is plotted and stored at 'files/image\_plot\_pc50.jpg'



The size of the image plot with 50 PCs is 25066 bytes ~ 24.5 KB

### ***Image Plot with 100 PCs***

The image is plotted and stored at 'files/image\_plot\_pc100.jpg'



The size of the image plot with 100 PCs is 25182 bytes ~ 24.6 KB

### ***Image Plot with 150 PCs***

The image is plotted and stored at 'files/image\_plot\_pc150.jpg'



The size of the image plot with 150 PCs is 22498 bytes ~ 22 KB

### ***Image Plot with 200 PCs***

The image is plotted and stored at 'files/image\_plot\_pc200.jpg'



The size of the image plot with 200 PCs is 15979 bytes ~ 15.6 KB

### **Summary**

Size of original image = 36.4 KB

Size of original image plot = 38.3 KB

<b>K-Means Clustering of image color</b>		<b>Principal Component Analysis of image</b>	
<i>Number of clusters <math>k</math></i>	<i>Size of image plot (KB)</i>	<i>number of PCs</i>	<i>Size of image plot (kB)</i>
5	29.4	25	22.5
10	30.8	50	24.5
15	31.2	100	24.6
20	31.1	150	22.0
25	30.8	200	15.6



## Conclusion

- The image has been color quantized by clustering the RGB colors using K-Means clustering. With between 15-20 clusters, we obtain a decent color quantized image of good quality.

The file size of the image with  $k=18$  is 31489 bytes ~ 30.8 KB



x

- The image has been compressed by dimensionality reduction using its principal components. With around 150 principal components, we obtain a decent compressed image of good quality. Statistically, 174 components give us a cumulative variance of 100%, so using 174 PCs we can fully reconstruct the image.

The file size of the image with  $pc=174$  is 18221 bytes ~ 17.8 KB which is a decent amount of compression from the original size.



## References

K-means: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/kmeans>  
PCA: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/prcomp>