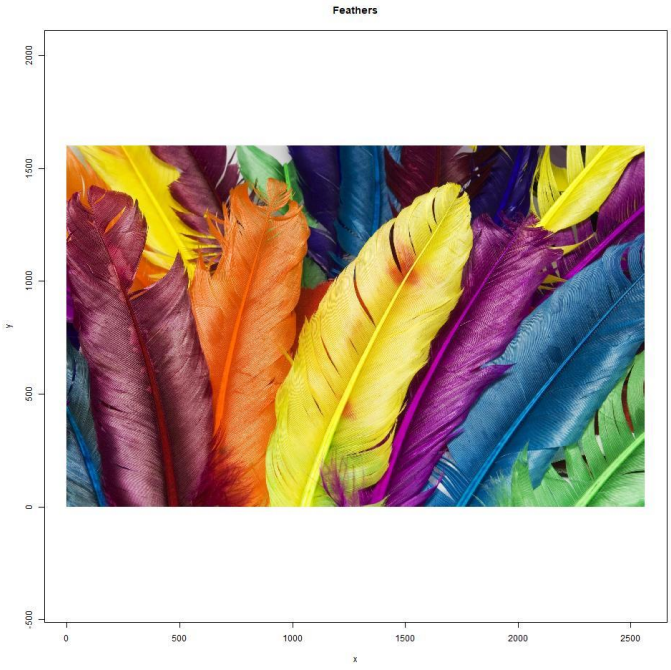
LAB ASSIGNMENT **LOKESWARI UMAKANTHAN (LXU190000)**

1. Dimensionality Reduction using K-Means

SNo.	Image	Size of the Image	K-value
1.	Original Image : **Dimension : (960*960*3)** 	125 KB	None
2.	**Dimension of all Images: (960*960*3)** 	118 KB	5

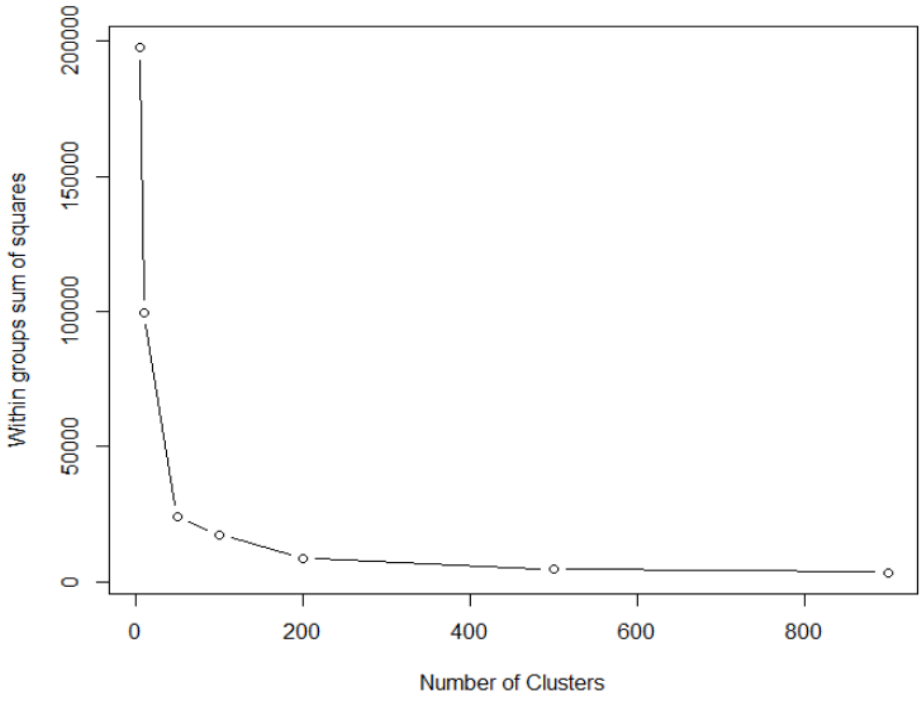
3.	Feathers  k-means cluster analysis of 10 colours	118 KB	10
4.	Feathers  k-means cluster analysis of 50 colours	119 KB	50

5.	Feathers  k-means cluster analysis of 100 colours	121 KB	100
6.	Feathers  k-means cluster analysis of 200 colours	118 KB	200

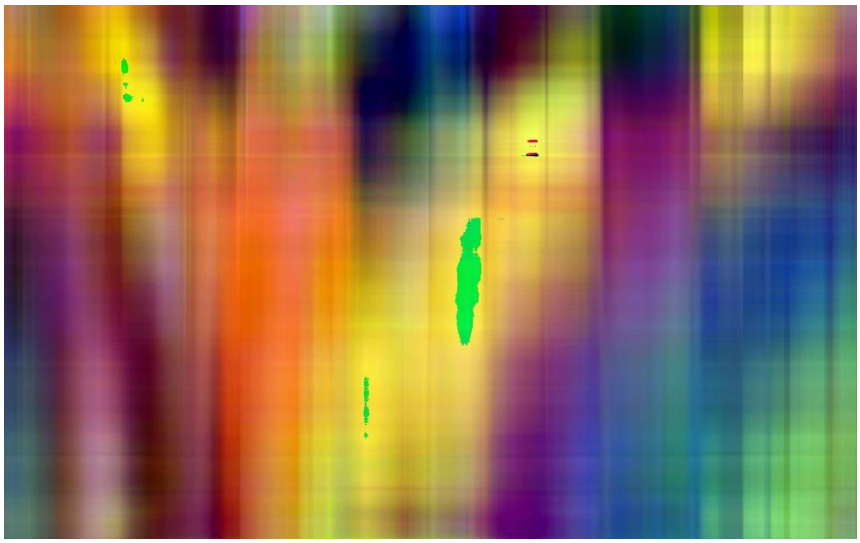
7.	Feathers  k-means cluster analysis of 500 colours	118 KB	500
8.	Feathers  k-means cluster analysis of 900 colours	117 KB	900

Best K-Means Value : **K =100**

- The Original Image has a size of 125 KB. The Dimensions of all the Images are saved as 960*960*3 for a larger view of the Image.
- The Image with k-value = 5 has reduced dimensionality attributes, hence the quality of the Image has greatly reduced with the significant loss in the information and color. At k=5, size = 118 KB.
- The Image with k-value = 10,50 also has significant information loss in the Image. As the k-value increases the Size of the Image also increases. At k=10, size = 118 KB and at k = 50, size = 119 KB.
- The Image with k-value = 100 has the image quality very similar to the original image. The loss of information is also very less compared to the previous images. The Image is not compressed so much and the size of the image is also nearly equal to the original image i.e. at k=100, size = 121KB. Thus this will be the best value for K.
- At k=200,500,900, the image clarity is good and similar to the original image but this takes a long time for the program to execute. The size of these images is also decreasing to 118KB,118KB and 117 KB respectively for k=200,500,900. Further increase in k will not have any significant effect on the image quality and the degree of image compression is also very less.
- From the Graph we infer that the best value for k =100. The **Elbow point** is the point at which K=100. Based on the we can conclude that the elbow point gives the best number of cluster in k-means algorithm graph.



2. Dimensionality Reduction using PCA

SNo.	Image	Size of the Image	PCA Components
1.	**Dimension of all Images : 2560*1600*3** 	157 KB	3
2.		586 KB	179

3.		674 KB	356
4.		697 KB	532
5.		697 KB	708

6.		693 KB	885
7.		684 KB	1061
8.		666 KB	1237

9.		645 KB	1414
10.		644 KB	1590

Best PCA component :

- When PCA components is 3, the images is so blurred and a lot of information loss can be noted. The Image compression is also very less, the size is reduced to 157 KB.
- When PCA components is 179, the image is resembling the original but at the same time misses few information. This needs some improvement to match the original image. The size of the image started to increase for PCA at 179 the size is 586 KB.
- When the PCA components is 365, 532, 708 are very similar to the original Image. Their image compression size if also increasing 674 KB, 697 KB, 697 KB respectively. At 708 as the number of PCA Components, the images looks very similar to the original Image with minimized information loss.

- This reaches a threshold at this point after this point the size of the image decreases with increase in PCA components.
- For the number of PCA components at 885, 1061, 1237, 1414, 1590, the size of the images decreases as 693 KB, 684 KB, 666 KB, 645 KB, 644 KB. The Improvement in the image quality is very small.

Compression Ratio for K-Means :

```
[1] "K_means_10.jpg size: 120.35 original: 128.439 % diff: -6%"
[1] "K_means_100.jpg size: 124.025 original: 128.439 % diff: -3%"
[1] "K_means_200.jpg size: 121.65 original: 128.439 % diff: -5%"
[1] "K_means_5.jpg size: 121.751 original: 128.439 % diff: -5%"
[1] "K_means_50.jpg size: 122.823 original: 128.439 % diff: -4%"
[1] "K_means_500.jpg size: 120.963 original: 128.439 % diff: -6%"
[1] "K_means_900.jpg size: 120.768 original: 128.439 % diff: -6%"
```

- ❖ The K-Means with number of cluster = 100 has reduced the image by 3% from 124 KB to 128 KB. Thus the difference in the file size converge to around 3% which cannot result in a further efficient compression ratio. Hence k = 100 is the efficient value.

Compression Ratio for PCA :

```
[1] "feathers_compressed_1061_components.jpg size: 701.163 original: 128.439 % diff: 446%"
[1] "feathers_compressed_1237_components.jpg size: 682.03 original: 128.439 % diff: 431%"
[1] "feathers_compressed_1414_components.jpg size: 661.392 original: 128.439 % diff: 415%"
[1] "feathers_compressed_1590_components.jpg size: 659.708 original: 128.439 % diff: 414%"
[1] "feathers_compressed_179_components.jpg size: 600.429 original: 128.439 % diff: 367%"
[1] "feathers_compressed_3_components.jpg size: 161.725 original: 128.439 % diff: 26%"
[1] "feathers_compressed_356_components.jpg size: 690.602 original: 128.439 % diff: 438%"
[1] "feathers_compressed_532_components.jpg size: 714.257 original: 128.439 % diff: 456%"
[1] "feathers_compressed_708_components.jpg size: 714.498 original: 128.439 % diff: 456%"
[1] "feathers_compressed_885_components.jpg size: 709.811 original: 128.439 % diff: 453%"
```

- ❖ The PCA components with the number as 708 has the highest compression ratio. The difference in the file size from 714 KB to 128 KB has a reduction percentage of 456%. Thus the highest image compression ratio with minimal loss of information is provided by the image produced with 708 PCA components.