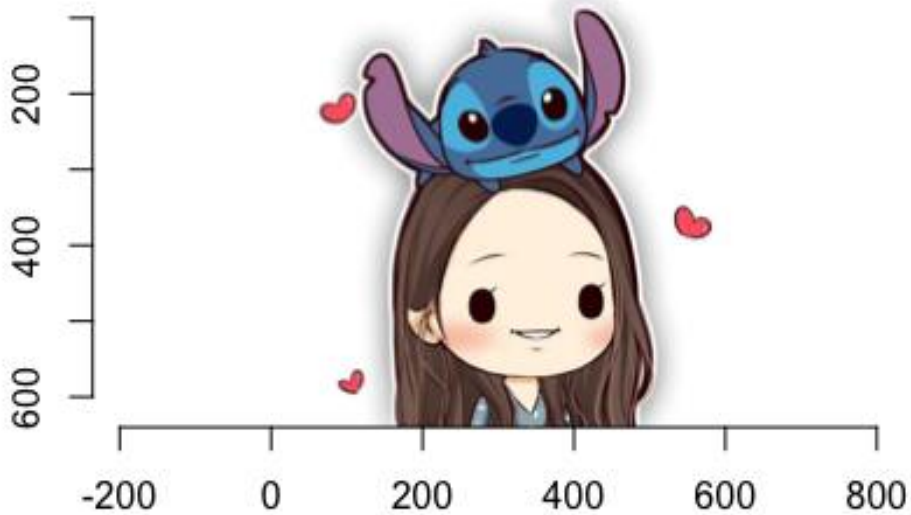# A_1

Jiayue Wu

01/11/2020

I start by setting a seed using my student number, which ensures that I will get the same result each time I run the same process.

```
set.seed(1004768165)
source("functions.R")
```

Here I read in the image. This is the image we will use and get its cross-stitch plot.

```
image_file <- "img.jpg"
im <- imager::load.image(image_file)
plot(im)
```
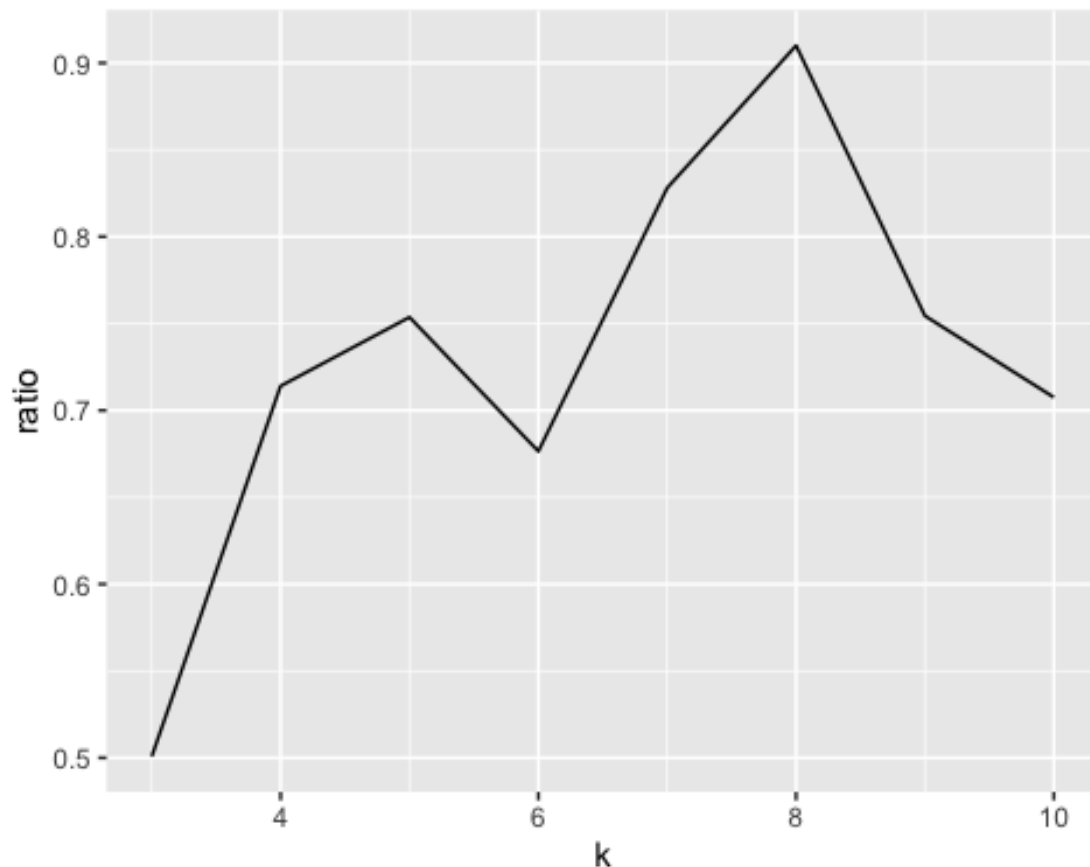
STEP 1

```
my_clusters <- process_image(image_file_name = image_file,k_list=2:10)
```

Here I perform a clustering using function process_image() with a list of k's. It shows a list of cluster information that we will use next for the image—image_file_name we uploaded.

STEP 2

```
scree_plot(cluster_info = my_clusters)
```


Then I use a scree plot using the scree_plot function, derived from my_clusters. More specifically, I use ratio plot here, i.e. total between SS by various k. The x axis is the value of k and the y axis is the average total between sum of squares, i.e. ratio k = $\frac{tot.ss[k]}{tot.ss[k-1]}$

We want the k that maintain more significant infomation instead of similar information compare the prvious k. What's more, I will prefer a cluster with relatively low ratio (has different informstion from previous cluster) and the next cluster with relatively large ratio(do not include relatively more information compare to the k cluster). This is obvious that k before 7 all have low ratio (no higher than 0.81). And k = 7 has very low ratio(around 0.66), which indicates that k = 7 has significant change in the information retained. While after 7, the ratio increases sharply，with ratio 0.775 when k = 8 and the plot reaches to the peak when k = 9, with ratio around 0.9. This also means we do not see a significant change in the information retained as we 9 clusters. Hence, I will consider k = 7 or k = 8 here.
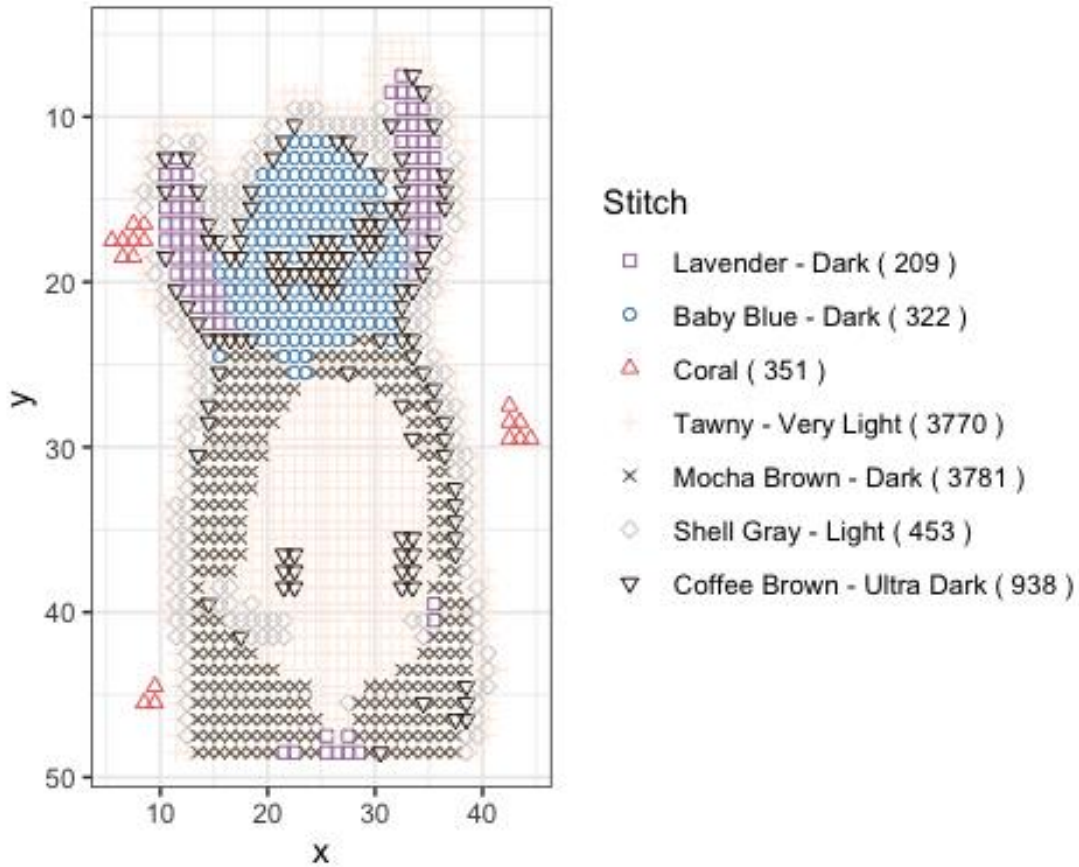
STEP 3

```
colour_strips(cluster_info = my_clusters)
```
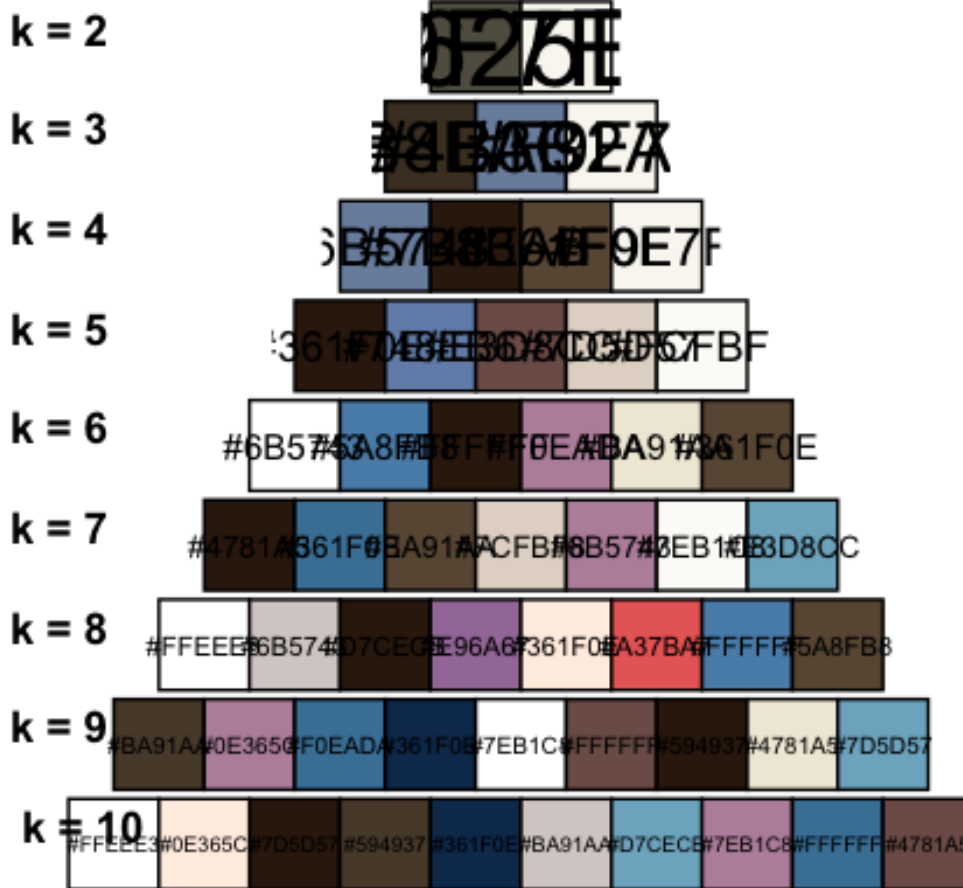

To specify our cluster number, I use the colour_strips function, which produces colour strips with the DMC colour closest to the cluster centre colour. Here, I will mainly compared the colour between k = 7 and k = 8 to decide which cluster I should choose. By observing, the colour strip with k = 8 has relatively large difference with colour strip with k = 7. While k = 8 has very similar colours with clusters go beyond 8. Hence, I will choose k = 8 here.

STEP 4

```
make_pattern(cluster_info = my_clusters, k=8, x_size = 50, black_white = FALSE, backg
round_colour = "B5200")
```

Stitch

□	Lavender - Dark (209)
○	Baby Blue - Dark (322)
△	Coral (351)
	Tawny - Very Light (3770)
×	Mocha Brown - Dark (3781)
◇	Shell Gray - Light (453)
▽	Coffee Brown - Ultra Dark (938)

Lastly, I created a 50 x 50 Stitch in colour cross-stitch using the make_pattern function. This is derived from my_cluster. k is the number of clusters I will use. Aforementioned, I will use k = 8. Here x_size means the (approximate) total number of possible stitches in the horizontal direction. Here I use x_size = 50. And I set black_white = FALSE to get a in colour cross-stitch. (We can also build a cross-stitch to be in black and white by changing the setting of "black_white = FALSE" to be "black_white = TRUE" or "black_white". I will remove the background_colour in this cross-stitch. By observing the original plot, I find that the background_colour is white, with dmc code = "B5200". I enter it as the background_colour to remove it.