Assignment 1 - Vignette

WS

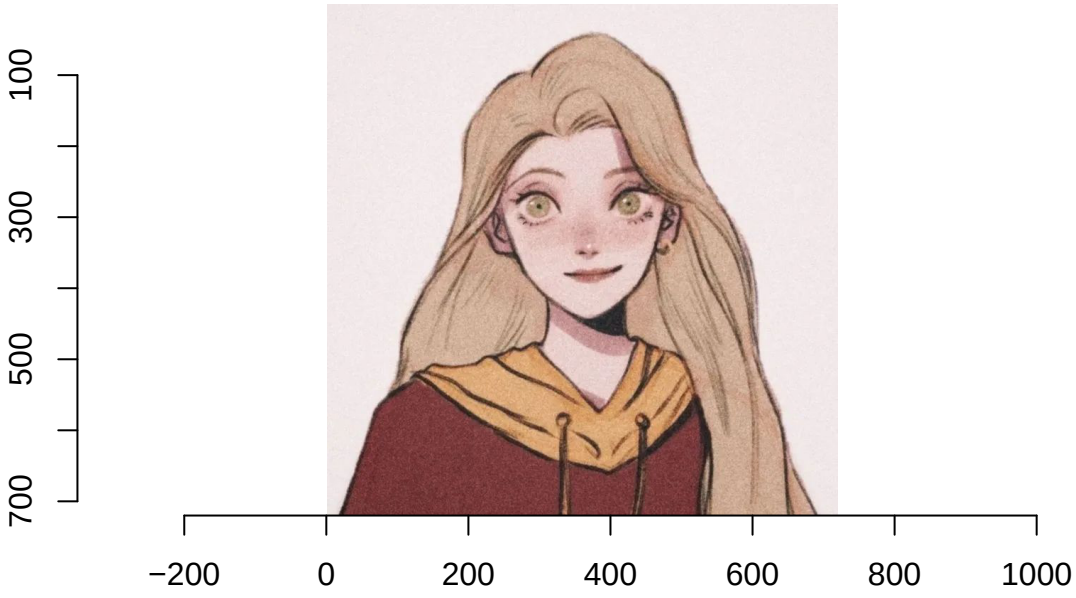
03/11/2020

Firstly, I set the seed to the last four digits of my student number, and attach the function file functions.R .

```
set.seed(8590)
source("functions.R")
```

Loading the image. I choose a image that has pure background and resolution of 720 × 720.

```
image <- "image10.png"
im <- imager::load.image(image)
plot(im)
```



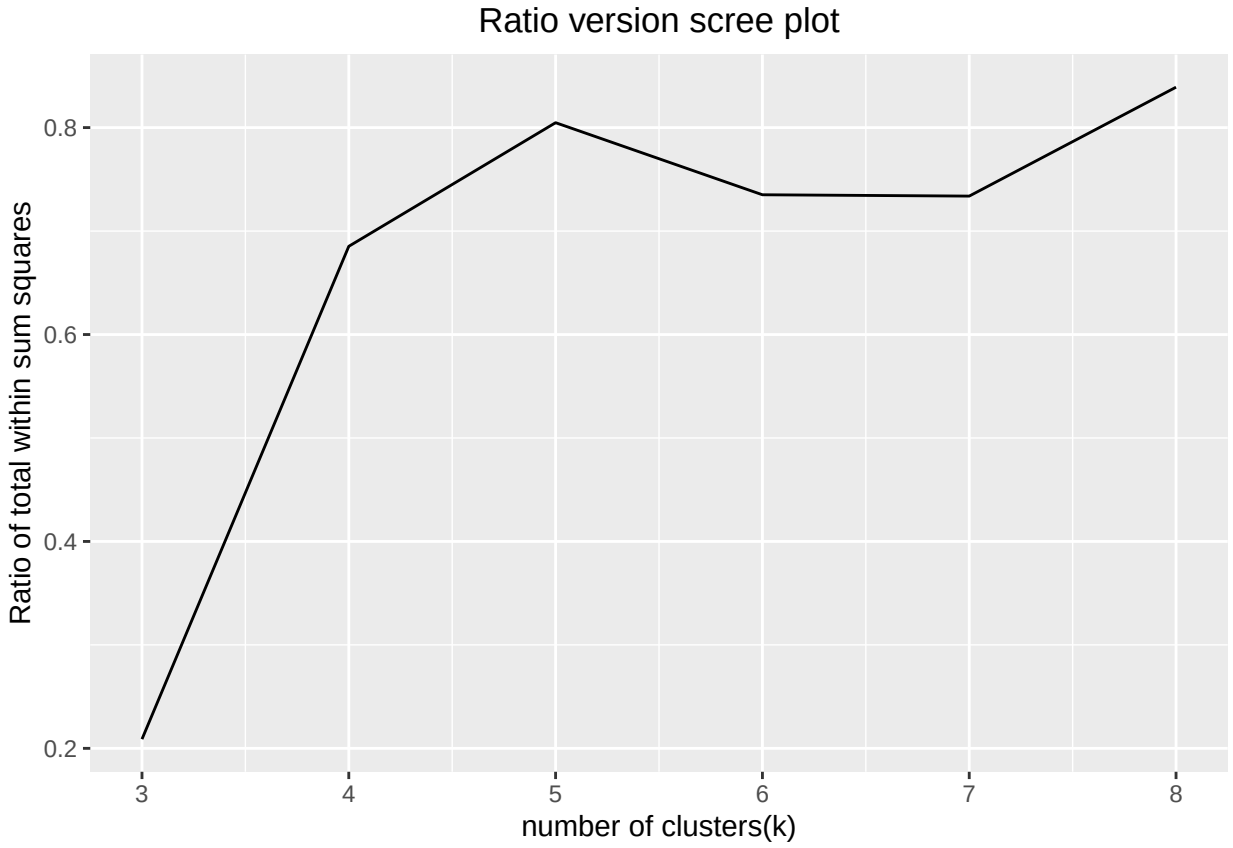
I am wondering how many clusters is good to present this image. Therefore I chose a range of of k values (where k states for the number of clusters), here I choose to compute information for two clusters, three clusters, ..., to eight clusters. To find the right number of clusters present this image best.
The following is the numerical data frame shows the information I computed for different number of clusters.

```
clusters <- process_image(image, k = 2:8)
clusters
```

```
## # A tibble: 7 x 8
##       k kclust    totss tot.withinss betweenss  iter centers     tidy_dat
##   <int> <list>    <dbl>        <dbl>     <dbl> <int> <list>      <list>
## 1     2 <kmeans> 83803.       24353.    59450.     1 <tibble [2~ <tibble [518,4~
## 2     3 <kmeans> 83803.        5087.    78716.     2 <tibble [3~ <tibble [518,4~
## 3     4 <kmeans> 83803.        3486.    80317.     3 <tibble [4~ <tibble [518,4~
## 4     5 <kmeans> 83803.        2805.    80998.     3 <tibble [5~ <tibble [518,4~
## 5     6 <kmeans> 83803.        2062.    81741.     3 <tibble [6~ <tibble [518,4~
## 6     7 <kmeans> 83803.        1513.    82290.     3 <tibble [7~ <tibble [518,4~
## 7     8 <kmeans> 83803.        1270.    82533.     7 <tibble [8~ <tibble [518,4~
```

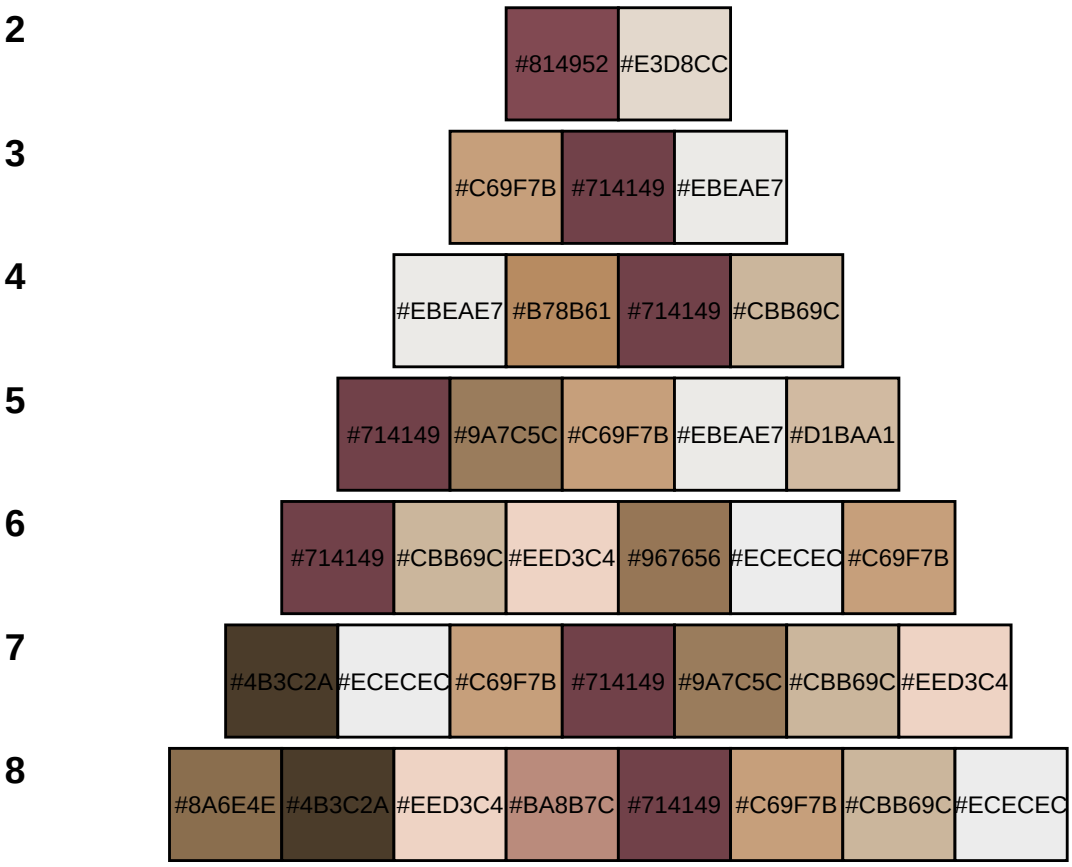
Then I plot a ratio version scree plot by using the above information I compute for two to eight clusters. Note the ratio when k = k_i is compute by total within sum square of k_i divides total within sum square of k_i-1.
A high ratio k = k_i means that the total within sum square from k_i-1 clusters to k_i clusters changes a lot. I can observed that when k equals to 5, I get a high ratio. Moreover, for numbers after 5, which is for k equals 6, 7, 8... the ratio does not change a lot, ratios approach steady.
Therefore, I will choose k = 4. Which is choose to divide to four clusters.

```
scree_plot(clusters)
```



Then I want to find the color information for each clusters.
This following color strips shows the DMC color of different clusters of different number of clusters. For example, when k = 3, there are three clusters, each cluster has a center. The DMC color for points in each cluster equals to the DMC color closets to their center color. Since the three cluster centers has the closet DMC color of "#C69F7B", "#714149", and "#EBEAE7", all the points are in those three colors when k = 3. I can observed that if the image contains four clusters, the four colors looks different to each other. However, when the image contains five clusters, some of the colors looks similar, for example "#9A7C5C", "#C69F7B", "#D1BAA1" looks similar. That means five clusters is not so good.
Therefore I will choose k = 4. Which is choose four clusters.

```
color_strips(clusters)
```



Produce a lower resolution image and plot the vignette pattern from it by using all the information I computed above. Plot the pattern with four clusters.

The following graph shows the vignette pattern with four colors. Moreover, people can set background_colour = "EBEAE7" to extract the background color.
Or people can set black_white = TRUE to make the pattern shows in black and white.

```
make_pattern(clusters, k = 4, x_size = 80, black_white = FALSE, background_colour = NULL)
```