STA314 A1

TONGFEI ZHOU

10/31/2020

For the assignment, I load the necessary packages here. However, in the function colour_strips, I will need the packages dplyr and ggplot2 since the extra help Professor provided uses them, and so I load them here as well. Please download any packages that is not on your computer.

```
library(imager)
library(tidyverse)
library(tidymodels)
library(sp)
library(scales)
library(cowplot)
devtools::install_github("sharlagelfand/dmc")
library(dmc)
library(dplyr)
library(ggplot2)
```

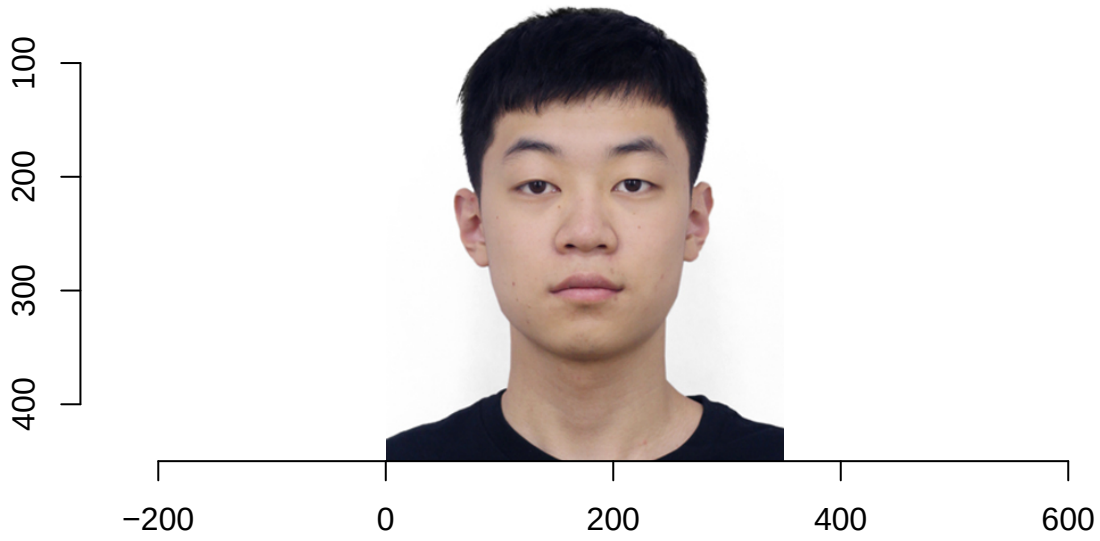
To be able to generate the same results every time, I will set the seed with my last 4 student number.

Furthermore, I call the functions.R file I have written here so that I can use the functions written there without copying the code anymore.

```
set.seed(8448) #last 4 digits of my student number
source("functions.R")
```

I first plot the image I will use, here I use my own self portrait. (So this is what I look like :)), and this requires the package imager.

```
im <- imager::load.image("self_portrait.jpg")
plot(im)
```



I run the process_image function to get all the necessary information about my self portratit picture. The function generally provides the kmeans information about my image, for further details, one can check the functions.R file for the documentation part.

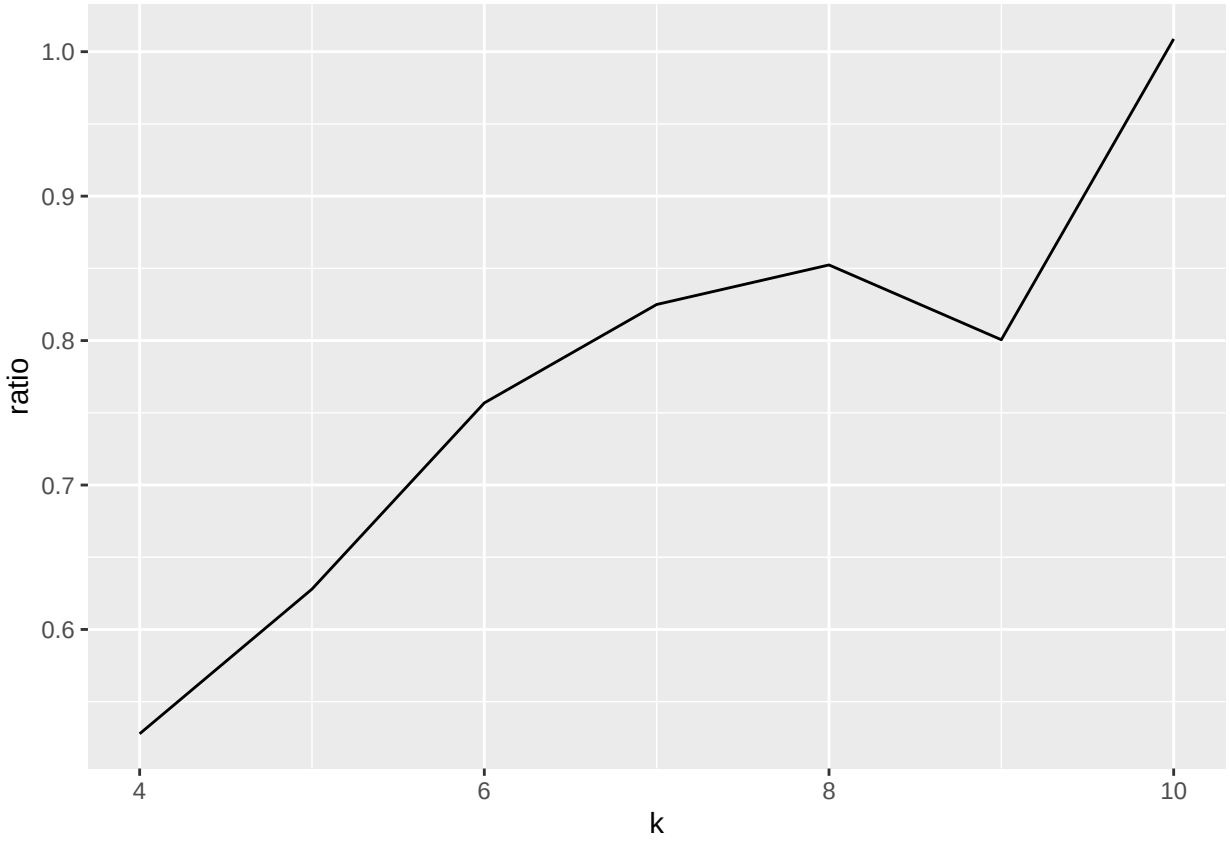
```
clusters <- process_image(image_file_name="self_portrait.jpg", k_list=3:10)
clusters
```

```
## # A tibble: 8 x 8
##       k kclust   totss tot.withinss betweenss  iter centres      tidy_dat
##   <int> <list>   <dbl>        <dbl>     <dbl> <int> <list>       <list>
## 1     3 <kmean~ 45431.        1555.    43876.     2 <tibble [3 ~ <tibble [157,5~
## 2     4 <kmean~ 45431.         821.    44610.     2 <tibble [4 ~ <tibble [157,5~
## 3     5 <kmean~ 45431.         515.    44916.     3 <tibble [5 ~ <tibble [157,5~
## 4     6 <kmean~ 45431.         390.    45041.     3 <tibble [6 ~ <tibble [157,5~
## 5     7 <kmean~ 45431.         322.    45109.     4 <tibble [7 ~ <tibble [157,5~
## 6     8 <kmean~ 45431.         274.    45157.     3 <tibble [8 ~ <tibble [157,5~
## 7     9 <kmean~ 45431.         220.    45212.     4 <tibble [9 ~ <tibble [157,5~
## 8    10 <kmean~ 45431.         222.    45210.     4 <tibble [10~ <tibble [157,5~
```

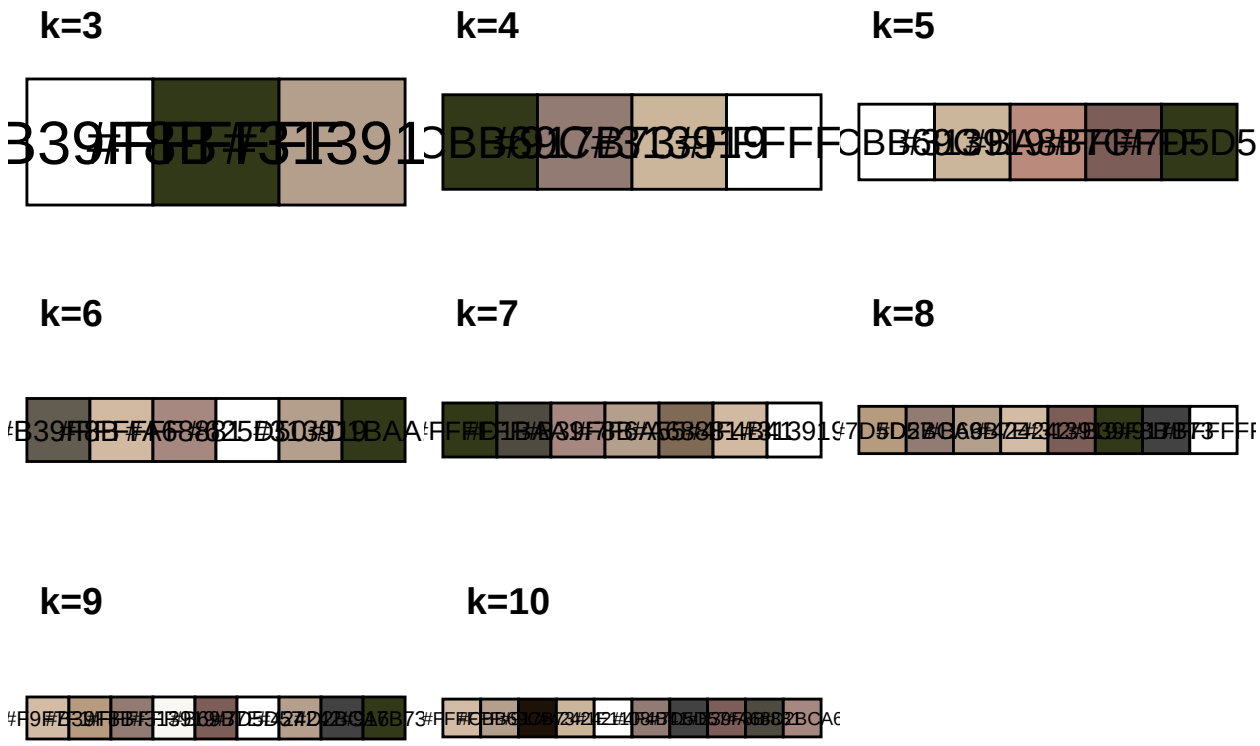
As one can see from above, it produces the information for each k, the tot.withinss is each k's cost function's value, and totss is the sum of them all. For the kclust, it provides how each point assigned to which cluster for each k, and also their mean point value and all information provides by R, but we have thousands of points, and so I will not print them here.

I produce the scree plot in the ratio version and also the colour strips to help me choose the best k I will use in the make_pattern function. Here the parameter in the function will be my clusters generated above, and for how I use it, check the functions.R file.

```
scree_plot(clusters)
```



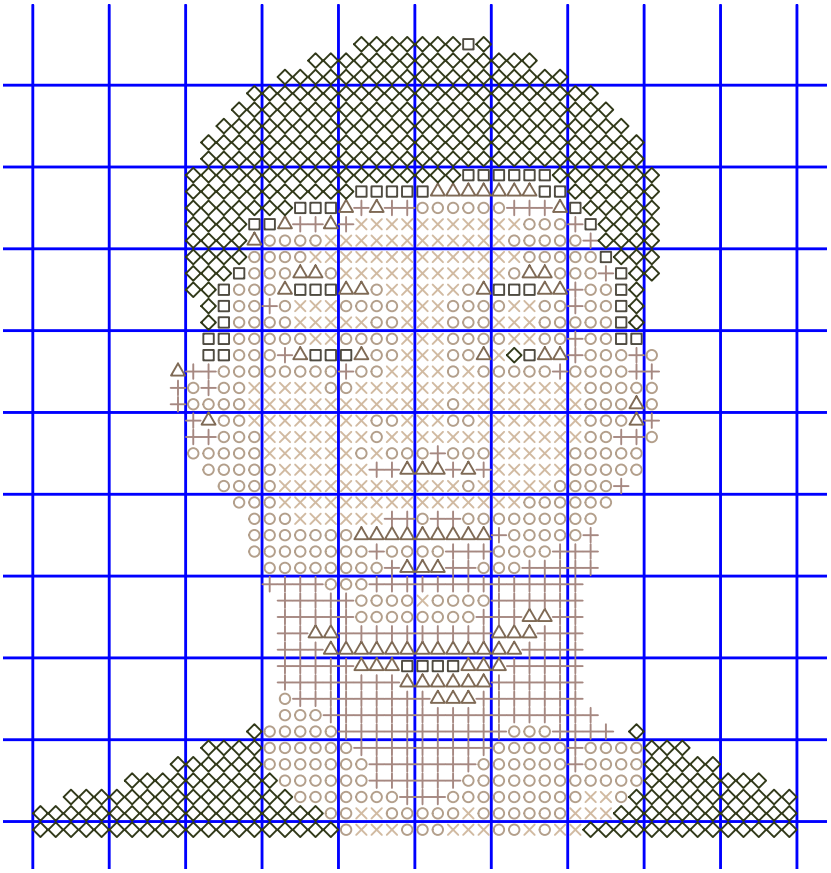

```
colour_strips(clusters)
```



From the ratio scree plot and the colour strips shown above, one can see that after $k = 7$, the ratio of $k = 8$ to $k = 7$ is almost 1 (0.85 from the graph), and also from the colour strips one can see that after $k = 7$, in all $k = 8, k = 9$ and $k = 10$ groups there are some similar colours, and in the last two groups there are some almost same colours, and therefore, we should choose $k = 7$ to reduce the size and represent all the necessary information.

Finally, I produce the cross-stich as required with $k = 7$. Notice that, even though here the background_colour parameter is set to be the default none, after the function plot the cross-stich, I identify that the Snow White is my image's background color, and therefore, I set the background_colour to be the dmccode of the Snow White, B5200, in the function make_pattern. As a consequence, what you see the final cross-stich here is the one without background colour and I set the parameter into "B5200".

```
make_pattern(cluster_info = clusters, k=7, x_size=50, black_white=FALSE, background_colour = "B5200")
```



So from the final cross-stich, we can see that for example, my hair and my clothes are assigned to same color, Avocado Green - Black, with dmccode 934, and my skin color is Medium Mocha Brown with dmccode 3032. The blue grid line is added to help read the pattern. Notice that the background colour Snow White does not show on the legend on the right.