

MTH208A - Data Science Lab I

Project Report

Color Variation Analysis of Paintings
of Some Famous Painters

By Group No: 3

- Khyati Singh (221332)
- Shubham Kumar Rajak (221423)
- Abhinav Kumar (221252)
- Rohit Dutta (221396)

Under the guidance of Dr. Dootika Vats

Acknowledgement

We group-3 would like to express our profound gratitude towards **Dr. Dootika Vats**, our academic and project instructor for MTH208A(Data Science Lab), for her guidance and constant supervision throughout the process and providing creative ideas and necessary information regarding the project which led to the completion of this project. It has been a great learning experience and this project has provided us with a practical insight of the theoretical knowledge gathered during the course lecture.

Introduction

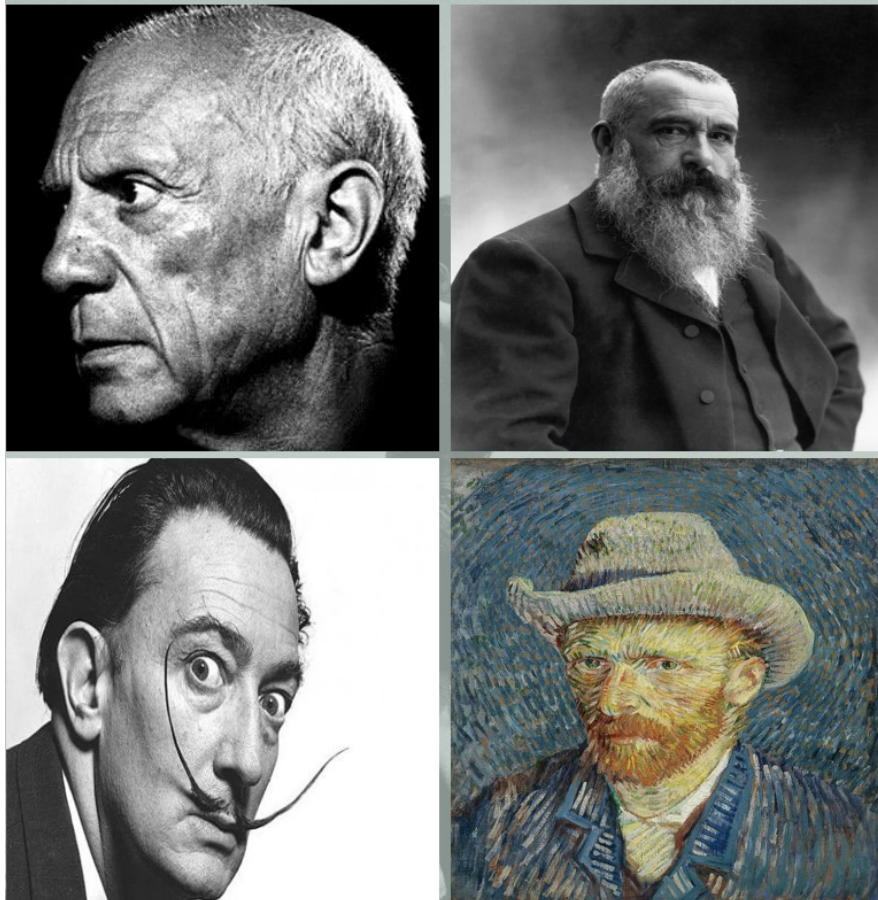
Colors are everywhere around us. And while the first thing we notice in a painting is its color, we do not realize that there is a lot more we are taking in. Color communicates so much. It has the power to convey moods and feelings like when people say, “*He was green with envy*”, “*I am feeling blue*” or “*She was red hot*”. It is for this reason that choosing colors and choosing where they fit in our lives is essential to how we live and the kind of lives we try to cultivate.

Now talking about paintings, “*Painting*” is a form of artistic expression that has been known to be practiced by mankind since the dawn of civilization. The world has seen some phenomenal painters and artists, some of the famous painters of the world who made their contribution to art include Leonardo da Vinci, Raphael, Claude Monet, Frida Kahlo, Salvador Dali, Vincent Van Gogh, Pablo Picasso, Michelangelo, Rembrandt to name a few. While the true definition of painting is subject to interpretation, in the end, it is the epitome of an artist’s emotions and creativity translated into a pattern of strokes and colors.

As beautifully quoted by someone ” ***Painting is the magical conjunction of space and no space, movement in stillness. It is a form of expressing emotions, depicting things we observe.***”

In works of art, artists use color to depict and describe the subject. Artists, especially painters, utilize their knowledge of color to portray mood, light, depth, and point of view in a work of art. Different painters used to follow different styles in their paintings. Some painters used to have less color variations, whereas some of them loved to use various colors within their paintings, which we tried to analyse here in this project. Our motive is to visualize different color palettes and color variations in various paintings of some world famous painters. By analyzing this, can we try to interpret what a painter is trying to depict to some extent.

For this study we have considered four painters namely Pablo Picasso, Claude Monet, Salvador Dali and Vincent Van Gogh.



Description of the Data:

- **Preliminary Data:** To analyse the color variations in paintings of different painters, paintings from different years were collected as potential data set. Corresponding to each painter, that is, Pablo Picasso, Claude Monet, Vincent Van Gogh and Salvador Dali, mainly name of the paintings, years they were made in and their corresponding links were scrapped from <https://www.wikiart.org/> and https://commons.wikimedia.org/wiki/Main_Page.
- **Obtaining the Preliminary Data:** All the information mentioned in the preliminary data were extracted from the mentioned websites via **web scrapping** through library *rvest* in R studio.
- **Obtaining the Final Data:** Corresponding to each painter a data frame consisting of name of the paintings, year in which they were made and the painting links are obtained. After that, to get relevant quantitative observations corresponding to each painting of the painters, *imager* library in R studio is used. First to check whether we can load and work with the images, *for loop* had been used to load the painting links by the function *load.image()* corresponding to all the painters. For some links, error occurred in loading the image because of incorrect standard format of dimensions of image and some of the images were not in *.jpg*, *.jpeg* or in *.png* format. In that cases we noted down the index of the painting in that data frame. Along with these indices we also noted down the indices of the images with size more than 5 mb and the paintings for which the year in which it was made was not available. After checking for all the paintings of a particular painter, we removed those rows from the data set having the mentioned indices. Next we found that the data set corresponding to Claude Monet had 714 rows, but the data sets corresponding to other painters had more than 714 rows in it. So, for better compatibility we considered 714 rows from each of the four data frames. Then corresponding to each painter, medians and quartile deviations of the pixel values for three color channels red, green and blue was computed for all the paintings. Moreover, the color proportion corresponding to sixteen basic colors have been computed for each of the painting. Also by averaging the color proportion of 38 dark shades we made a variable denoting average dark shade in a particular painting. Similarly by averaging color proportions of 47 medium shades and 53 light shades we obtain two another variable denoting the average medium dark shade and average light dark shade used in a particular painting for a given painter.

Then finally we combined all the four data frames in a single data frame which we considered as our final data, named *final_data_grp3*, based on which we have worked in this project.

- **Relevant Features of the Data:**
 - There are 714 data points for each painters.
 - To compute the color proportion of a particular color in a painting, we have considered the proportion of pixels with which the given color has distance less than 0.5 unit (threshold value).
 - The sixteen basic color and all the colors with which dark, medium and light shade are constructed are taken from the link: https://www.rapidtables.com/web/color/RGB_Color.html
 - The final data frame namely *final_data_grp3*, has 2856 rows and 29 columns.

A Glimpse of the Data set (Showing the first 6 rows):

##	Painter	Painting_Names	Year	Links
## 1	Oscar_Claude_Monet	Monet Veduta di Rouelles	1858	https://upload.wikimedia.org/wikipedia/commons/f/fb/Monet_Veduta_di_Rouelles.jpg
## 2	Oscar_Claude_Monet	Monet-faisan-musée de rouen	1861	https://upload.wikimedia.org/wikipedia/commons/f/fd/Monet-faisan-mus%C3%A9e_de_rouen.jpg
## 3	Oscar_Claude_Monet	Coint d'atelier	1861	
## 4	Oscar_Claude_Monet	Trophies-of-the-hunt-Monet	1862	
## 5	Oscar_Claude_Monet	Monet - Wildenstein 1996, 15.png	1862	
## 6	Oscar_Claude_Monet	Monet - Cour de ferme en Normandie, 1863	1863	

```

## 3 https://upload.wikimedia.org/wikipedia/commons/8/88/Coint_d%27atelier.jpg
## 4 https://upload.wikimedia.org/wikipedia/commons/9/9b/Trophies-of-the-hunt-Monet.jpg
## 5 https://upload.wikimedia.org/wikipedia/commons/1/11/Monet_-_Wildenstein_1996%2C_15.png
## 6 https://upload.wikimedia.org/wikipedia/commons/4/48/Monet_-_Cour_de_ferme_en_Normandie%2C_1863.jpg
## Red_Median Green_Median Blue_Median Red_QD Green_QD Blue_QD
## 1 0.5019608 0.5098039 0.4117647 0.33921569 0.33333333 0.41372549
## 2 0.1254902 0.1137255 0.1019608 0.01960784 0.01764706 0.01176471
## 3 0.3529412 0.2549020 0.2745098 0.17647059 0.14313725 0.10196078
## 4 0.3333333 0.1490196 0.1098039 0.29019608 0.22352941 0.15294118
## 5 0.2666667 0.2156863 0.1568627 0.06666667 0.05098039 0.02941176
## 6 0.4235294 0.3686275 0.1686275 0.25098039 0.24705882 0.22941176
## Black White Red Lime Blue Yellow Cyan Magenta
## 1 0.3675201 0.430348315 0.00000000 0 0 0.000000e+00 0 0
## 2 0.9533080 0.001482412 0.00000000 0 0 0.000000e+00 0 0
## 3 0.4488134 0.046976034 0.03076128 0 0 1.032895e-02 0 0
## 4 0.5801882 0.003501961 0.00352549 0 0 0.000000e+00 0 0
## 5 0.8730037 0.002692308 0.00000000 0 0 0.000000e+00 0 0
## 6 0.3565363 0.246012972 0.00000000 0 0 3.899294e-05 0 0
## Silver Gray Maroon Olive Green Purple Teal
## 1 0.49874639 0.29620546 0.4220770 0.3112263 0.4493258 0.09892456 0.131019262
## 2 0.02003271 0.06836016 0.9688975 0.1478505 0.9614331 0.08926298 0.031360275
## 3 0.21964286 0.63641682 0.6198896 0.5894455 0.4432636 0.57359023 0.364849624
## 4 0.25520392 0.44220392 0.6800196 0.4403294 0.4079725 0.28956863 0.001011765
## 5 0.01154457 0.47335775 0.9633822 0.7887057 0.9193651 0.65648352 0.175177045
## 6 0.30776046 0.53399751 0.6273076 0.7007898 0.5404595 0.24337228 0.020280663
## Navy dark_shade_prop medium_shade_prop light_shade_prop
## 1 0.2623628 0.2376577 0.12664988 0.425435446
## 2 0.9417911 0.4095900 0.01656932 0.006998972
## 3 0.4743633 0.4558281 0.16889488 0.110017999
## 4 0.3759137 0.3192749 0.17189746 0.102406955
## 5 0.5843346 0.5417240 0.03494389 0.006780588
## 6 0.1182678 0.3628758 0.13914922 0.259655147

```

Possible Biases in the Data:

- The websites from where we have collected the data contains almost all the paintings of each of the painters. But all the paintings uploaded in the websites are clicked photographs. So, the original color shades used in the paintings may appear different to some extent in the photographs which will lead to biases in the data.
- The threshold we used to calculate different color proportions in a painting is subjective in nature. So, there is a possibility that some other threshold value will give more accurate data. Also for different color shades different thresholds may be appropriate. Hence for the choice of threshold value as 0.5 biases may arise.
- Here we considered limited number of color shades to compute different color proportions. So, for example there is a possibility that the yellow shade Van Gogh used is not close to the pure yellow shade we considered with RGB values (0,1,1). This will result bias in the data to some extent.
- The data includes some pencil sketches of the painters which will result high proportion of white and black colors in those paintings. But actually no color have used there. As a result this will lead to some biases in the data.

Interesting Questions to be Asked from the Data:

1. For a given painting can we produce a graph that tells the possible colors used in that painting, so that looking at the graph only, can we predict the possible colors used in the corresponding painting?
2. Can the extent of color variations for a given painting be measured?
3. In given painting what is the approximate color proportion of a particular color?
4. Whether on an average there were any differences in color variations in a painting with respect to different painters?
5. Whether there are any painter who on an average used to use multiple colors in his paintings and whether there are any painter who on an average used to use flat colors throughout the painting?
6. How the use of different colors had changed over the respective careers of the painters?
7. Given a painting, can it be determined that which painter it belongs to?

Important Visualizations:

To analyze the color variation of individual paintings of each painter, density plots and box plots of three color channels red, green and blue have been plotted. Also the bubble plot of different color proportions have been plotted. Some insights of these plots are given below corresponding to each painter.

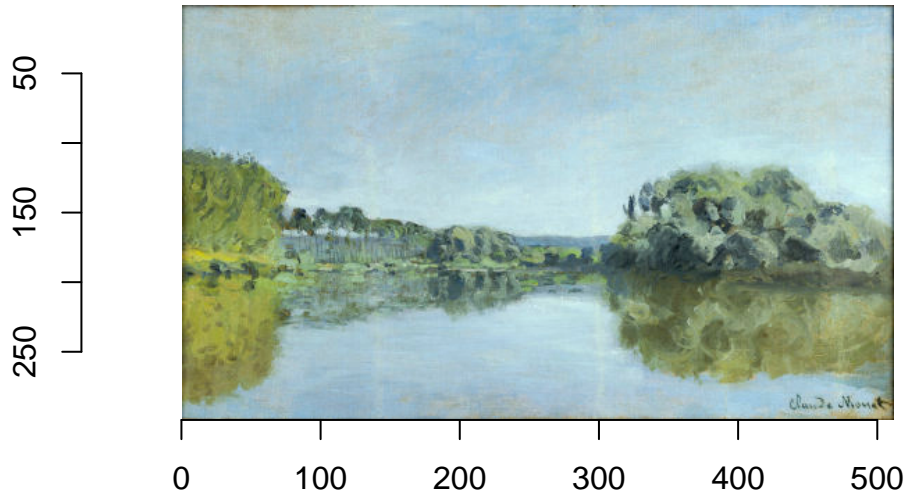
Note : Each pixel of an image consists of values corresponding to three color channels red, green and blue. Each color channel can take values from 0 to 255. In R studio, the values of color channels are represented by a proper fraction. That is, if $0 \leq x \leq 255$ be the pixel value corresponding to red color channel then the value in R studio will be $a/255$.

Pixel : A pixel is the smallest unit of a digital image or graphic that can be displayed and represented on a digital display device.

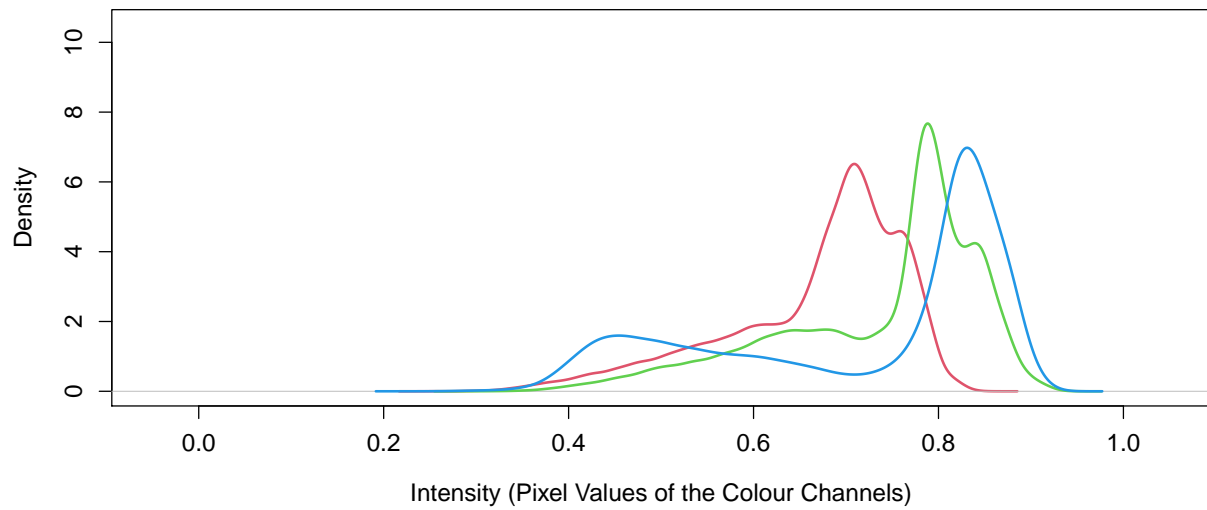
Claude Monet

Let us consider the painting given below, drawn in 1871.

Bords de la Seine a Argenteuil – Monet



Density Plot of Pixel Values of Three Color Channels Red, Green and Blue

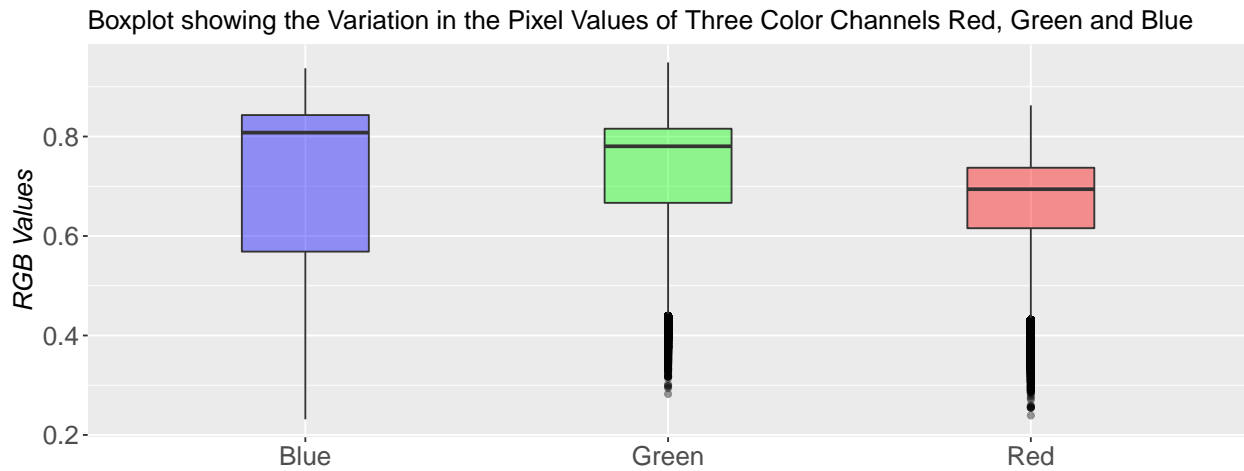


Observations:

From the density plot of three color channels it is observed that the red, green and blue color channels attain its peaks at around 0.71, 0.79 and 0.85. This means on an average in most of the pixels of the image have RGB value as **(0.71, 0.79, 0.85)**, which when multiplied by 255 gives the value of the triplet as **(183, 202, 217)** whose corresponding color shade is given below,

RGB
(183,202,217)

This above color is dominant in the painting.

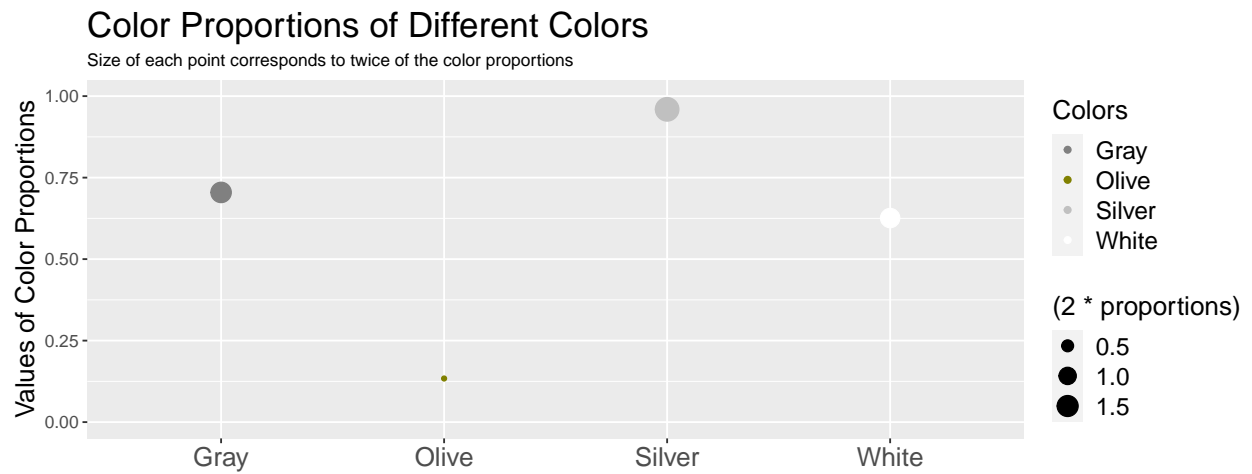


Observations:

From the box plots of three different color channels it is observed that box plot of blue is wider than that of red and green color channels. So the values corresponding to blue color channel is more dispersed than that of red and green. We can roughly say, the pixel values corresponding to red color channel is varying between 0.60 (equivalent to 153) to 0.72 (equivalent to 182). Similarly the range for green is in between 0.65 to 0.8 (equivalent to 165 to 204) and blue is in between 0.56 to 0.83 (equivalent to 142 to 211). Most of the pixels will take values covered by the Inter Quartile Range that is the width of the box plots. Now, some different RGB color values within the intervals mentioned above with respect to the given painting are given by,

RGB	RGB	RGB	RGB
(183,202,150)	(183,202,205)	(170, 190, 200)	(160, 180, 170)

Here all these shades are dominant in the painting we considered.



This is bubble plot of different color proportions where size of each of the bubble is $(2 * proportion)$.

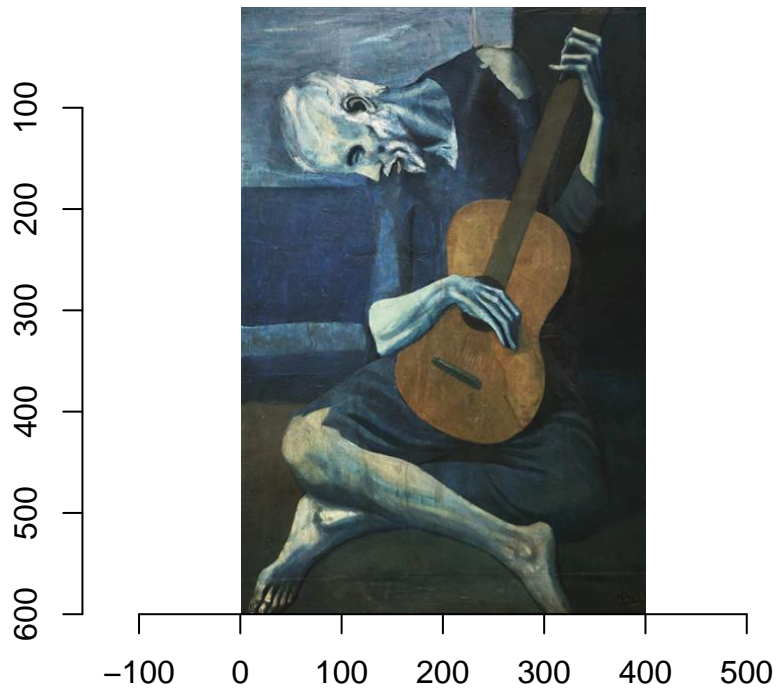
Observations:

From the above plot it is clearly seen that the color proportions corresponding to Gray, Silver and White are high, which is also visible in the original painting.

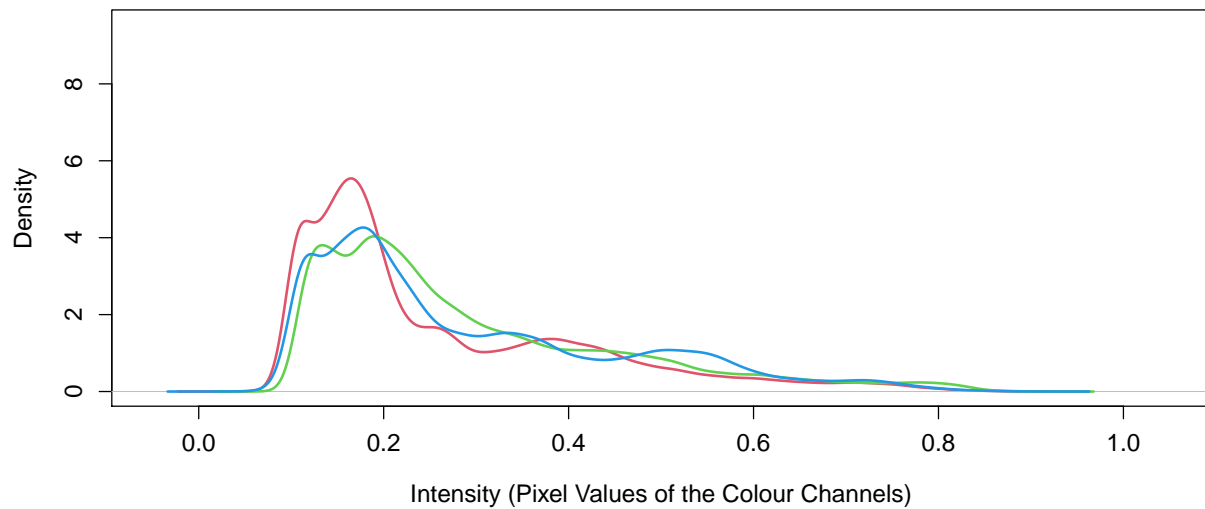
Pablo Picasso

Let us consider the painting given below, drawn in 1903.

The old blind guitarist

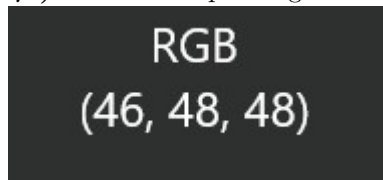


Density Plot of Pixel Values of Three Color Channels Red, Green and Blue

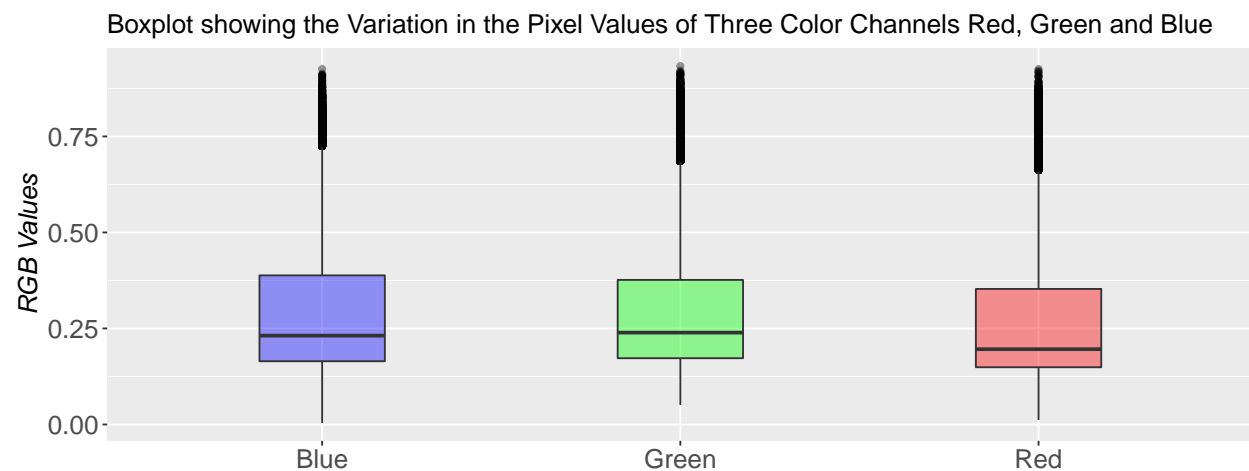


Observations:

From the density plot of three color channels it is observed that the red, green and blue color channels attain its peaks at around 0.71, 0.79 and 0.85. This means on an average in most of the pixels of the image have RGB value as $(0.18, 0.19, 0.19)$, which when multiplied by 255 gives the value of the triplet as $(46, 48, 48)$ whose corresponding color shade is given below,



This above color is dominant in the painting.

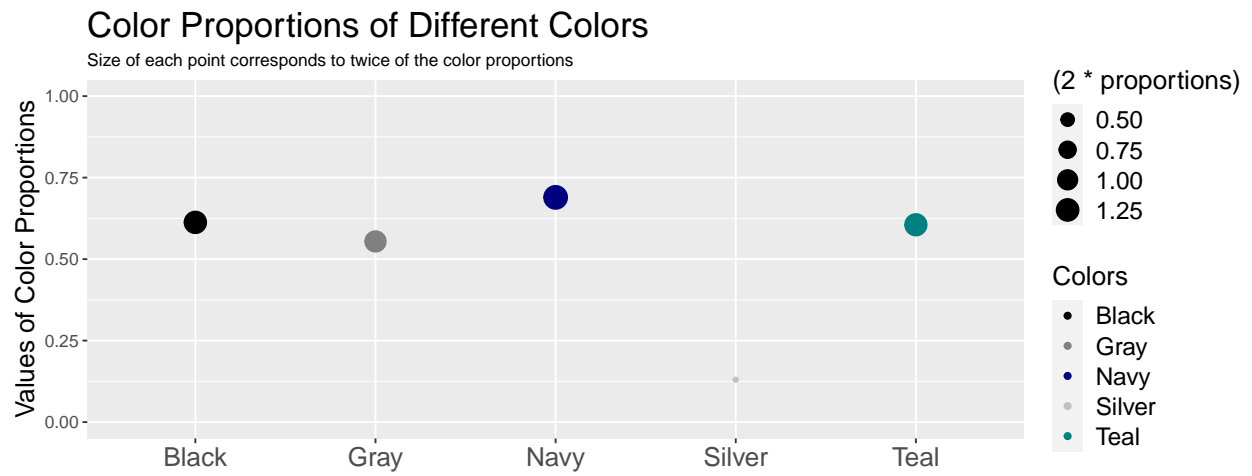


Observations:

From the box plots of three different color channels it is observed that box plot of blue, red and green color channels have almost same height also the median lines(horizontal black lines in the box plot) are almost co-linear. This indicates the image does not have too much variety of shades of different colors. We can roughly say, the pixel values corresponding to red color channel is varying between 0.15 (equivalent to 38) to 0.35 (equivalent to 89). Similarly the range for green is in between 0.17 to 0.38 (equivalent to 43 to 97) and blue is in between 0.16 to 0.39 (equivalent to 41 to 99). Most of the pixels will take values covered by the Inter Quartile Range that is the width of the box plots. Now, some different rgb color values within the intervals mentioned above with respect to the given painting are given by,



Here all these shades are dominant in the painting we considered.



This is bubble plot of different color proportions where size of each of the bubble is $(2 * proportion)$.

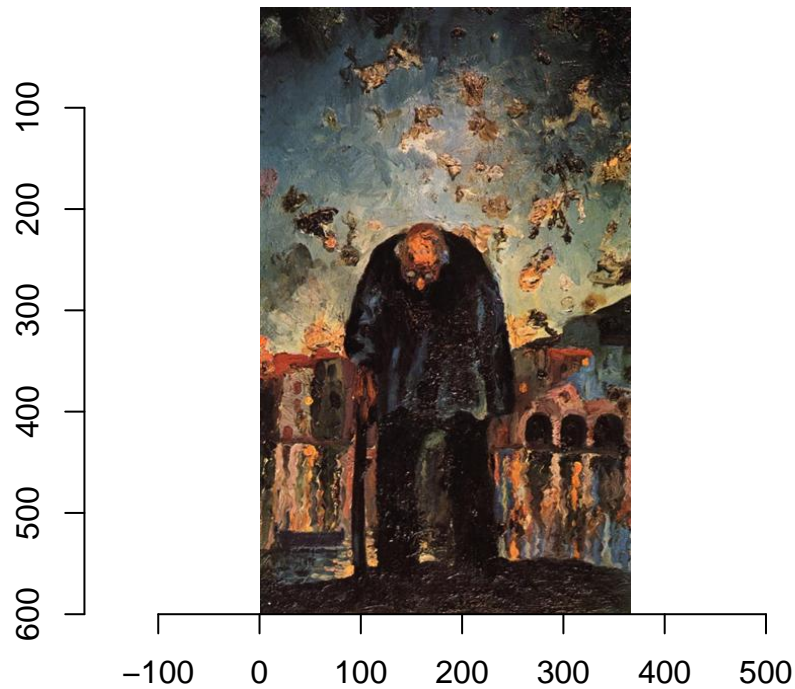
Observations:

From the above plot it is clearly seen that the color proportions corresponding to Gray, Black, Navy Blue and Teal are high, which is also visible in the original painting.

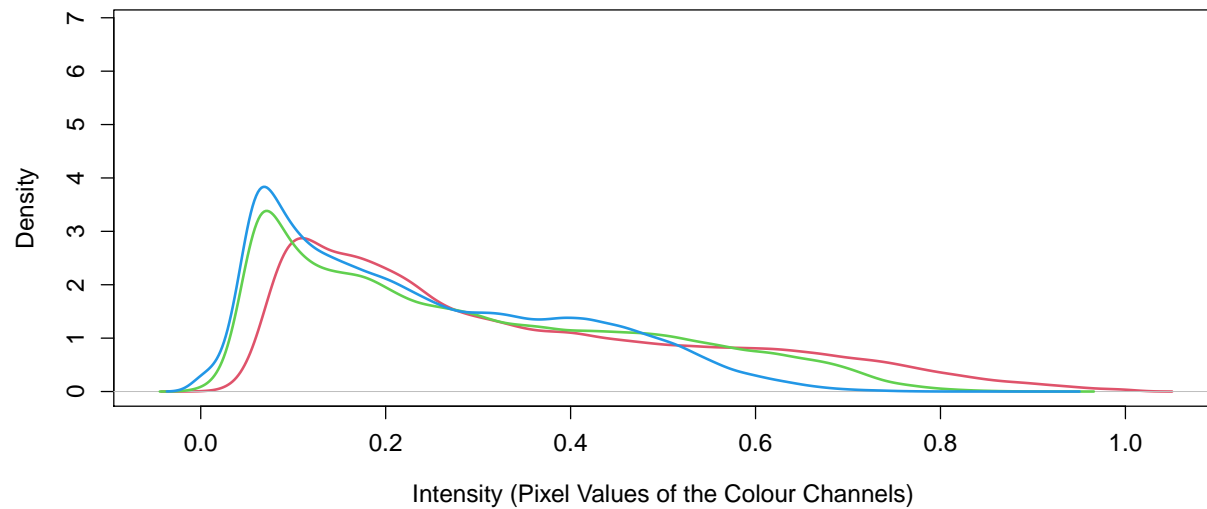
Salvador Dali

Let us consider the painting given below, drawn in 1918.

Crepuscular Old Man

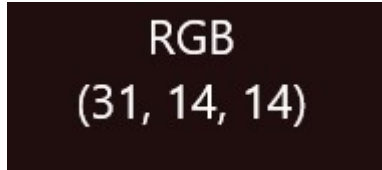


Density Plot of Pixel Values of Three Color Channels Red, Green and Blue

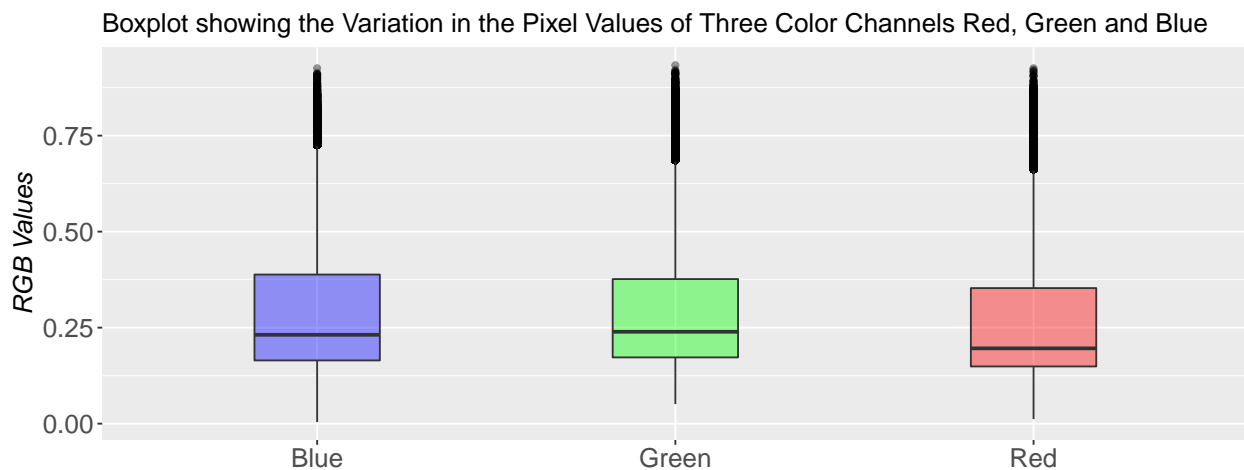


Observations:

From the density plot of three color channels it is observed that the red, green and blue color channels attain its peaks at around 0.71, 0.79 and 0.85. This means on an average in most of the pixels of the image have RGB value as $(0.12, 0.56, 0.056)$, which when multiplied by 255 gives the value of the triplet as $(31, 14, 14)$ whose corresponding color shade is given below,

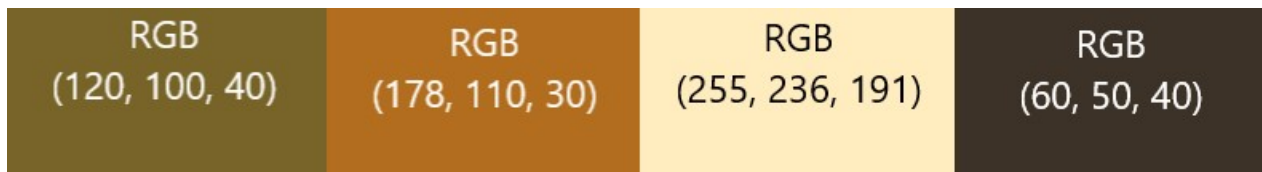


This above color is dominant in the painting.

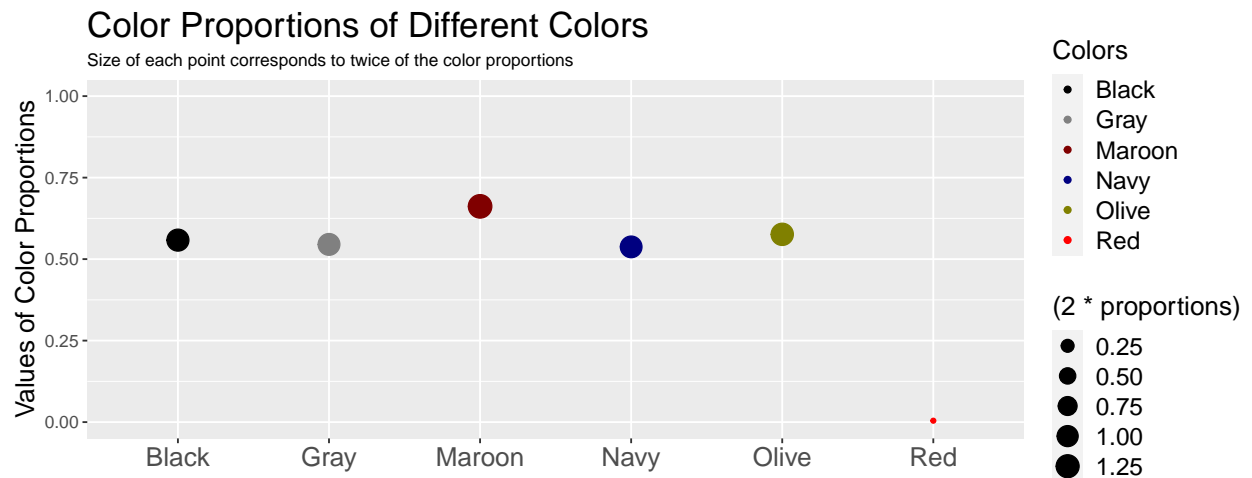


Observations:

From the box plots of three different color channels it is observed that box plot of red and green color channels have almost same height and are slightly greater than that of blue box plot. We can roughly say, the pixel values corresponding to red color channel is varying between 0.175 (equivalent to 45) to 0.50 (equivalent to 128). Similarly the range for green is in between 0.125 to 0.45 (equivalent to 32 to 115) and blue is in between 0.1 to 0.35 (equivalent to 25 to 89). Most of the pixels will take values covered by the Inter Quartile Range that is the width of the box plots. Moreover, densities of red and green color channels are comparatively higher than that of blue in the range 0.6 to 0.8 (equivalently from 153 to 204). So, we can possibly say that in some pixels red and green color channels will take value in the range 153 to 204, whereas blue takes value in the range 25 to 89, which results yellowish white shade in the pixels. Now, some different rgb color values within the intervals mentioned above with respect to the given painting are given by,



Here all these shades are dominant in the painting we considered.



This is bubble plot of different color proportions where size of each of the bubble is $(2 * proportion)$.

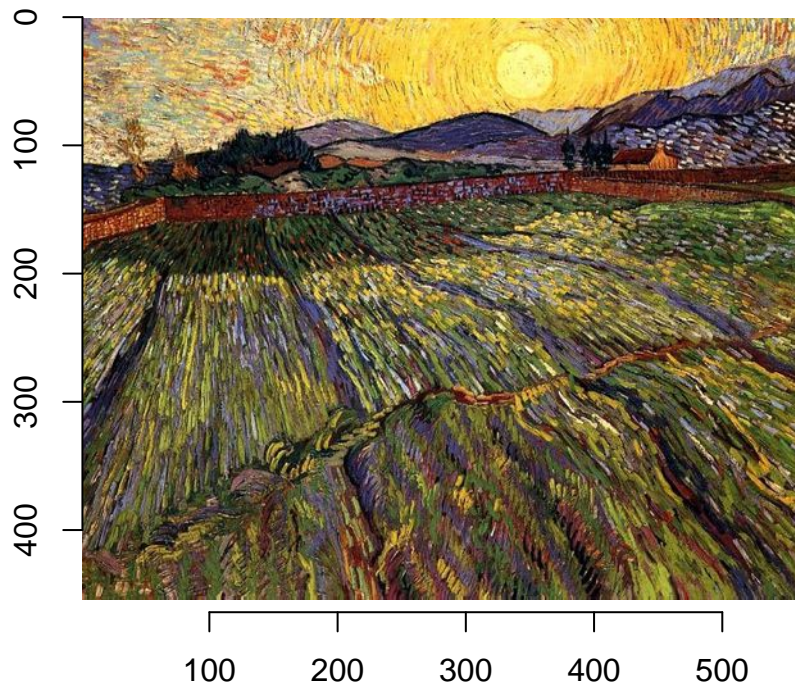
Observations:

From the above plot it is clearly seen that the color proportions corresponding to Gray, Black, Navy Blue, Maroon and Olive shades are high, which is also visible in the original painting.

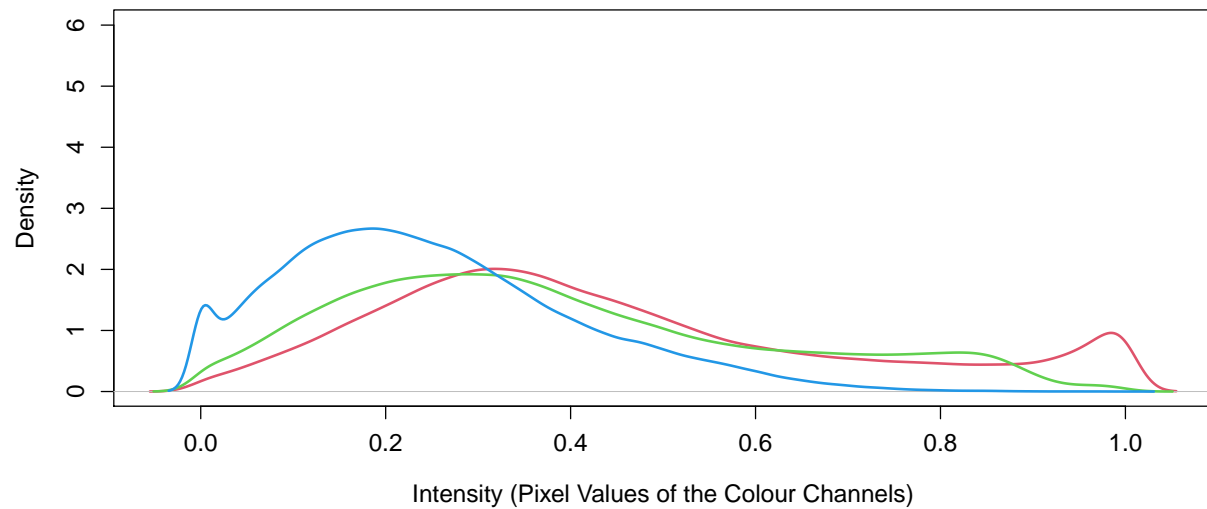
Van Gogh

Let us consider the painting given below, drawn in 1889.

Wheat Field with Rising Sun

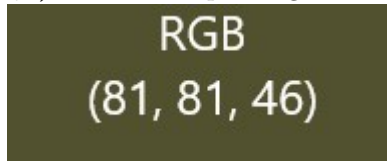


Density Plot of Pixel Values of Three Color Channels Red, Green and Blue

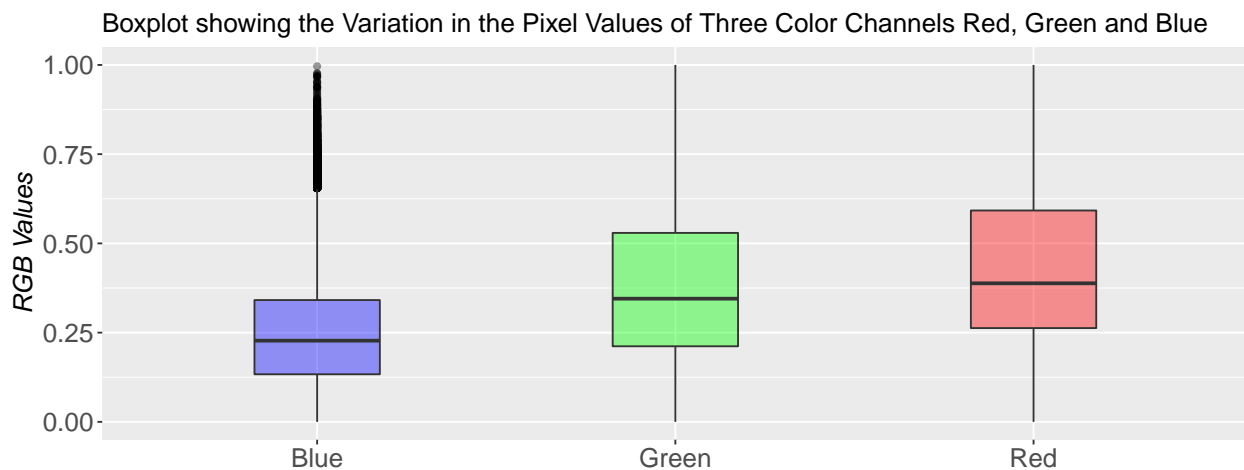


Observations:

From the density plot of three color channels it is observed that the red, green and blue color channels attain its peaks at around 0.71, 0.79 and 0.85. This means on an average in most of the pixels of the image have RGB value as $(0.32, 0.32, 0.18)$, which when multiplied by 255 gives the value of the triplet as $(81, 81, 46)$ whose corresponding color shade is given below,



This above color is dominant in the painting.

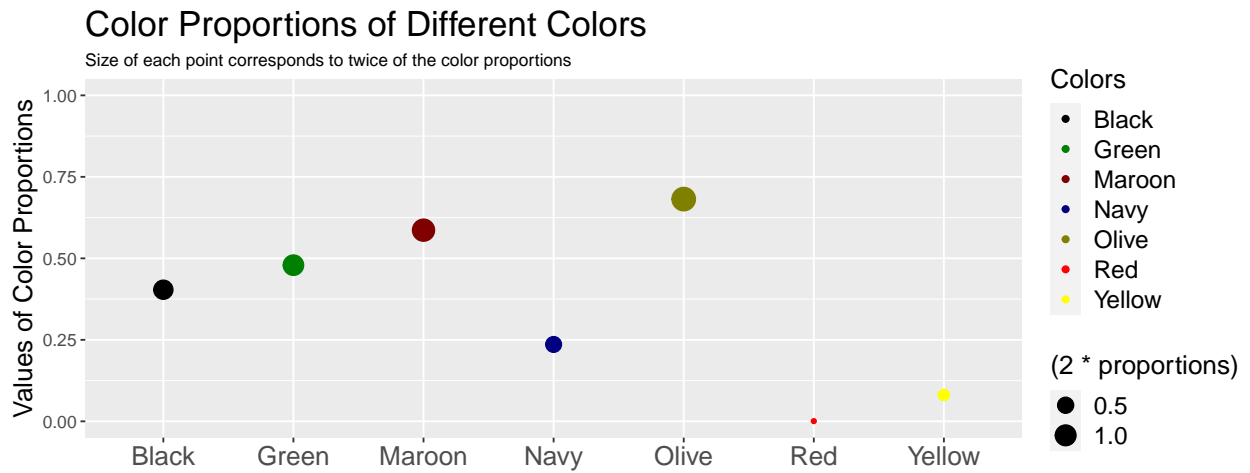


Observations:

From the box plots of three different color channels it is observed that box plot of red and green color channels have almost same height and are greater than that of blue box plot. We can roughly say, the pixel values corresponding to red color channel is varying between 0.25 (equivalent to 63.7) to 0.6 (equivalent to 153). Similarly the range for green is in between 0.2 to 0.525 (equivalent to 51 to 134) and blue is in between 0.125 to 0.35 (equivalent to 32 to 89). Most of the pixels will take values covered by the Inter Quartile Range that is the width of the box plots. Moreover, densities of red and green color channels are comparatively higher than that of blue in the range 0.6 to 0.9 (equivalently from 153 to 229). So, we can possibly say that in some pixels red and green color channels will take value in the range 153 to 229, whereas blue takes value in the range 32 to 89, which results yellowish white shade in the pixels. Again, from the density plot we can see that density of red color channel around 1 (equivalent to 255) is moderately high as compared to green and blue, moreover green and blue has some density around 0. So, there is a possibility that in some pixels the value of red color channel will dominate the others to a large extent, resulting reddish shade in the image. Now, some different rgb color values within the intervals mentioned above with respect to the given painting are given by,

RGB (227, 227, 30)	RGB (102, 82, 153)	RGB (80, 90, 60)	RGB (140, 70, 60)
-----------------------	-----------------------	---------------------	----------------------

Here all these shades are dominant in the painting we considered.



This is bubble plot of different color proportions where size of each of the bubble is $(2 * proportion)$.

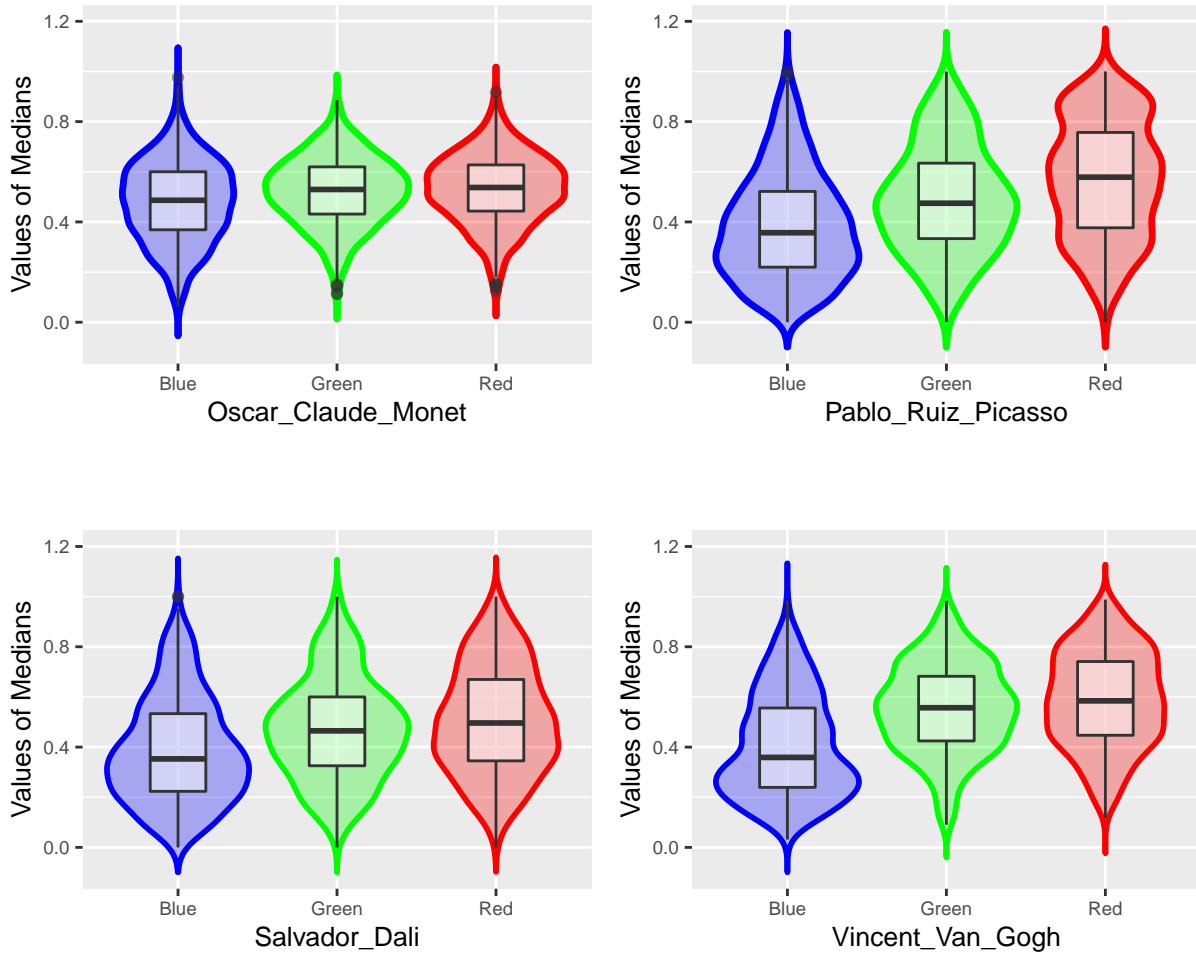
Observations:

From the above plot it is clearly seen that the color proportions corresponding to Olive, Black, Navy Blue, Maroon, Green, Yellow shades are high, which is also visible in the original painting.

Box Plots and Violin plots for Comparative Study:

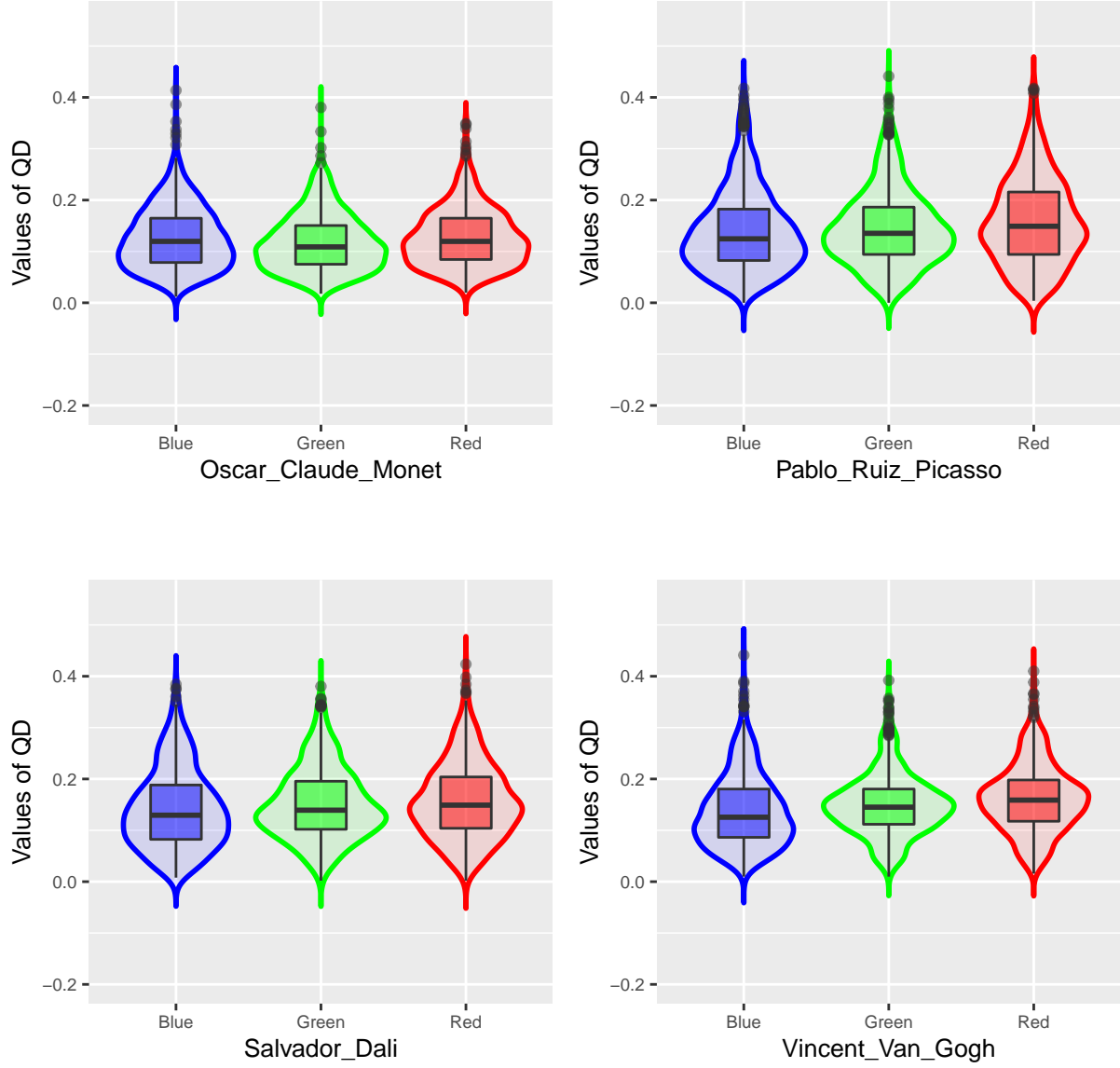
To analyse how on an average variety of color shades were used in the paintings corresponding to the four painters, box plots and violin plots of medians and quartile deviations of different color channels are plotted by considering all the paintings in our data set corresponding to each painters.

Box Plot and Violin Plots of Medians of Three Color Channels Considering All the Images of the Painters



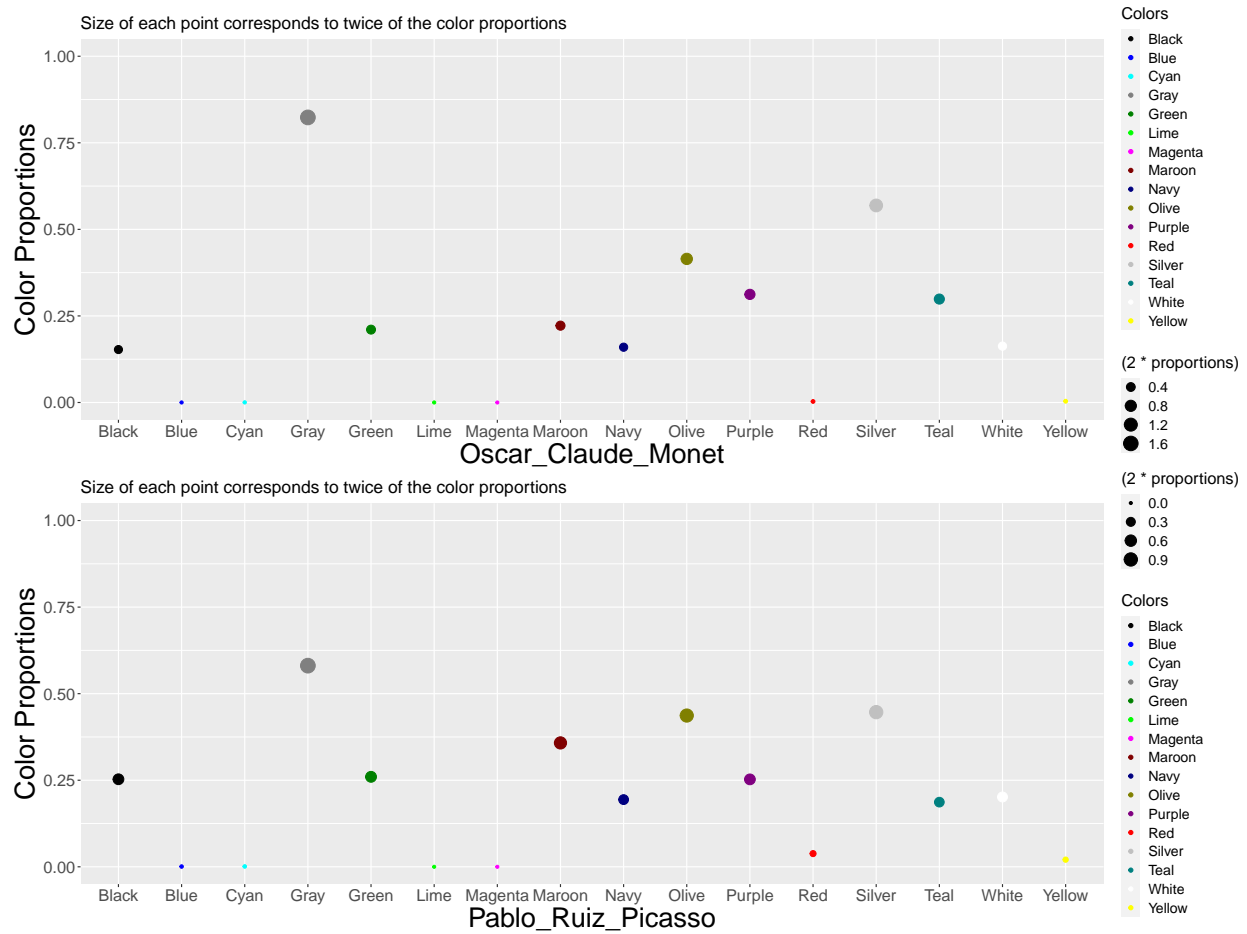
From the plot it is seen that the width of box plots of medians corresponding to Claude Monet is relatively smaller than the width of the box plots of the three color channels corresponding to Pablo Picasso, Salvador Dali and Van Gogh. Also, for Claude Monet, we can observe the global peaks by looking at the violin plots corresponding to the three color channels red, green and blue. This possibly indicates that on an average in most of the paintings most of the pixel took RGB values at around $(0.5, 0.5, 0.5)$ with less dispersion. That is we can possibly say that Claude Monet used to use less color variations in his paintings, which can be verified by looking at the pictures of Monet that on an average most of the paintings had flat colors. But in case of Picasso, Salvador Dali and Van Gogh, the peaks of the distribution of medians corresponding to three color channels do not form at around similar RGB values. Which possibly indicates that these painters had used more color variations in their most of the paintings.

Box Plot and Violin Plots of Quartile Deviations of Three Color Channels Considering All the Images of the Painters



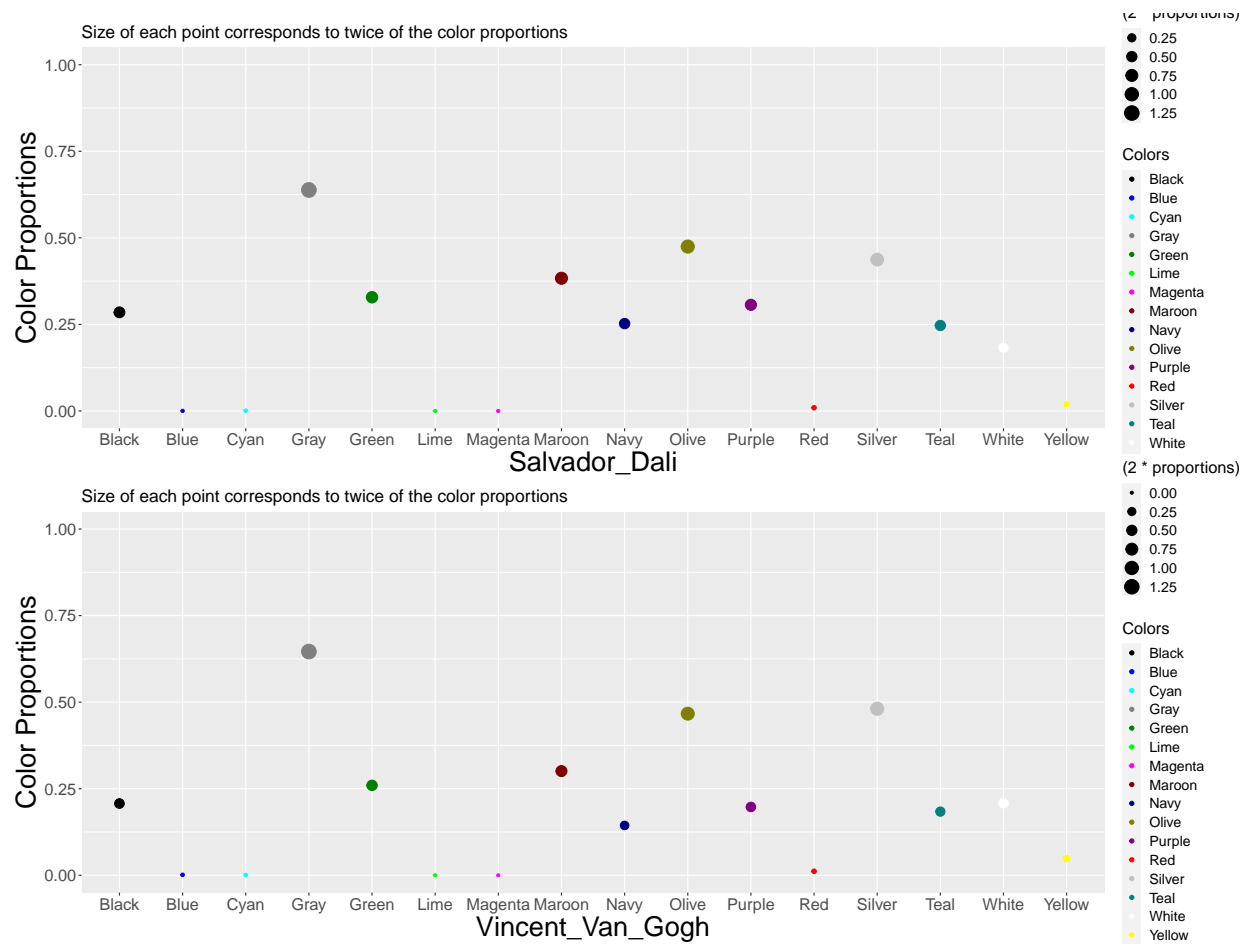
The box plots and violin plots of quartile deviations corresponding to three color channels indicate that on an average how the variations of the pixel values corresponding to three color channels used to vary. From the above graph we see that the width of box plots corresponding to Claude Monet and Van Gogh are comparatively lesser than that of Pablo Picasso and Salvador Dali. Moreover, the width of box plot of quartile deviations of red and blue color channels corresponding to Picasso and Salvador Dali was higher than the others, which possibly indicates that the variations of the pixel values corresponding to red and blue color channels used to vary a lot from painting to painting.

Bubble Plot of Average Color Proportions Corresponding to Different Painters



Above graph shows the average color proportion of the 16 base colors we have provided of all the paintings corresponding to Claude Monet and Pablo Picasso. The size of the bubble corresponds to twice of the color proportion. As we can see from the graph that the proportion of gray is high as compared to other colors, thus we can possibly say that it was dominantly used in Monet's paintings.

But the graph also shows that the average proportion in which red, blue and yellow was very low that is these are least used by Monet and Picasso, which is not true as we can see from the data that these colors was by Monet in a good proportion. This bias occurred because the shade of red, yellow used by Monet was different from the shade of red we have considered to calculate the proportions.

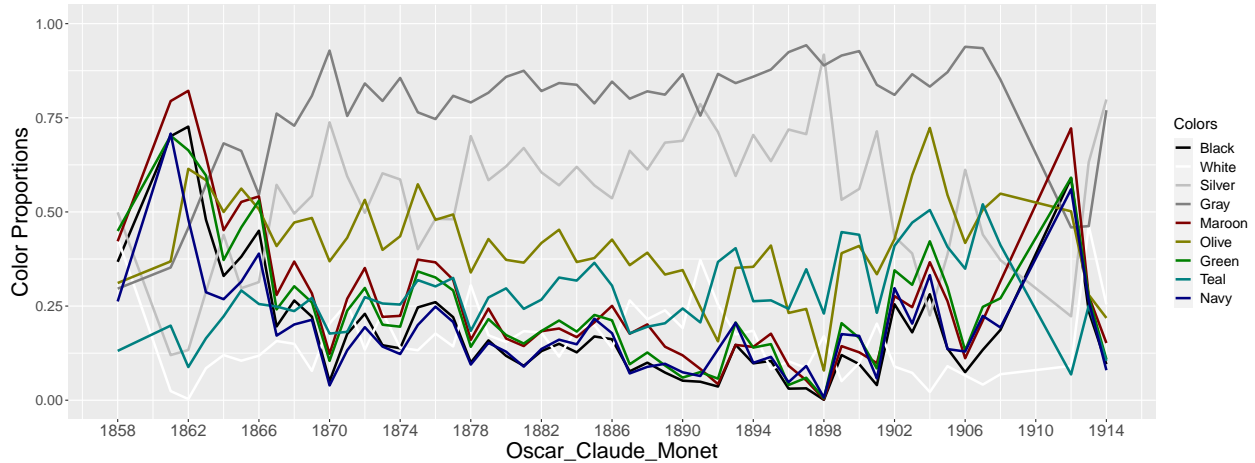


Above graph shows the average color proportion of the 16 base colors we have provided of all the paintings corresponding to Salvador Dali and Van Gogh. The size of the bubble corresponds to twice of the color proportion.

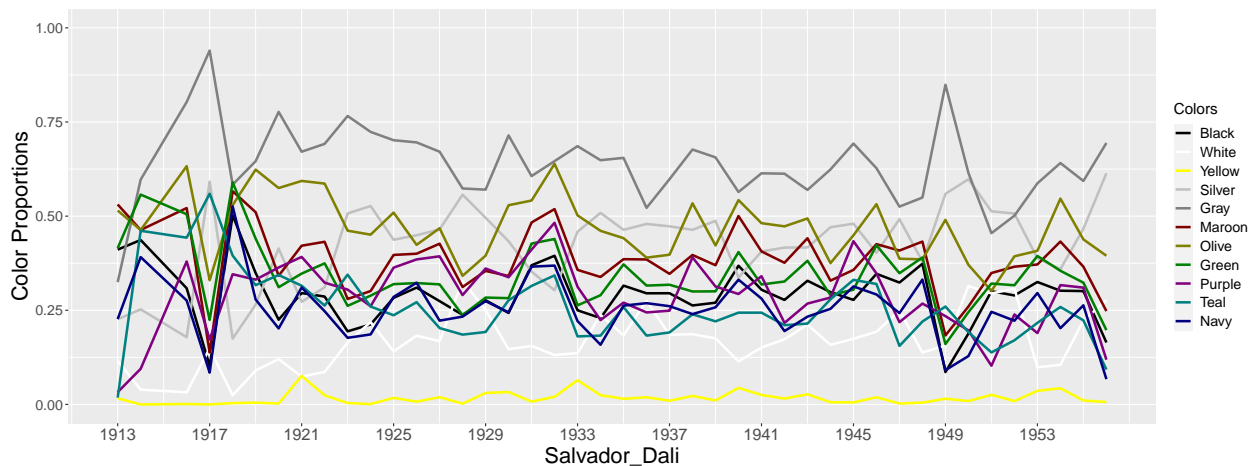
But the graph also shows that the average proportion in which red, blue and yellow was very low that is these are least used by Salvador Dali and Van Gogh, which is not true as we can see from the data that these colors was by Monet in a good proportion. This bias occurred because the shade of red, yellow used by Dali or Van Gogh was different from the shades we have considered to calculate the proportions. Moreover, it is seen that Van Gogh used to apply lots of yellow and blue shades in his paintings.

Line Diagrams of Average Color Proportions over the Years Corresponding to Different Painters

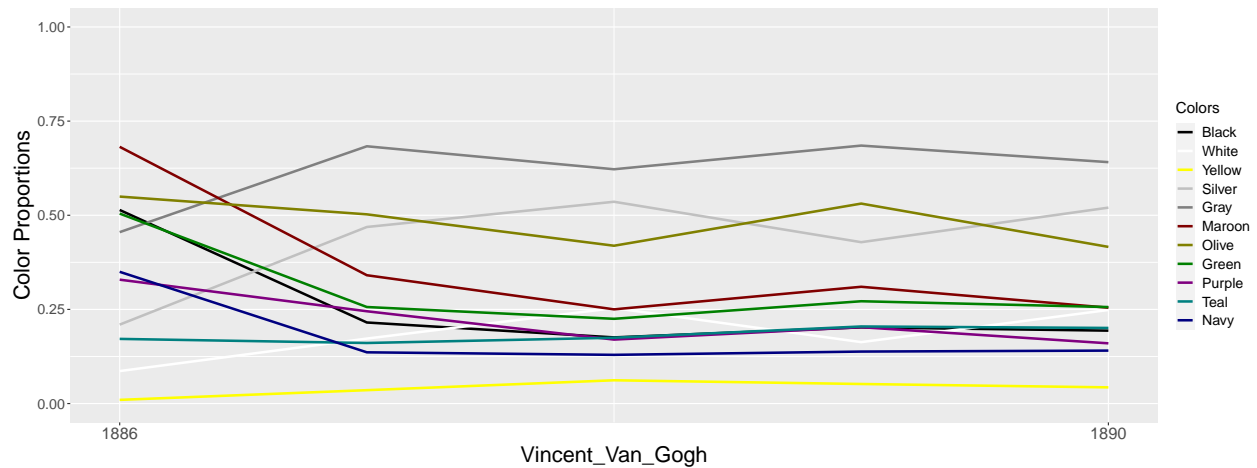
On plotting the values gained by taking the average of color proportions of all the paintings with respect to the basic colors for each year corresponding to a given painter, we get the following graphs.



Based on the data collected for Claude Monet(1858 - 1914),we can say from the above plot that the proportion of gray, silver and olive was higher as compared to other basic colors we have considered throughout the years.

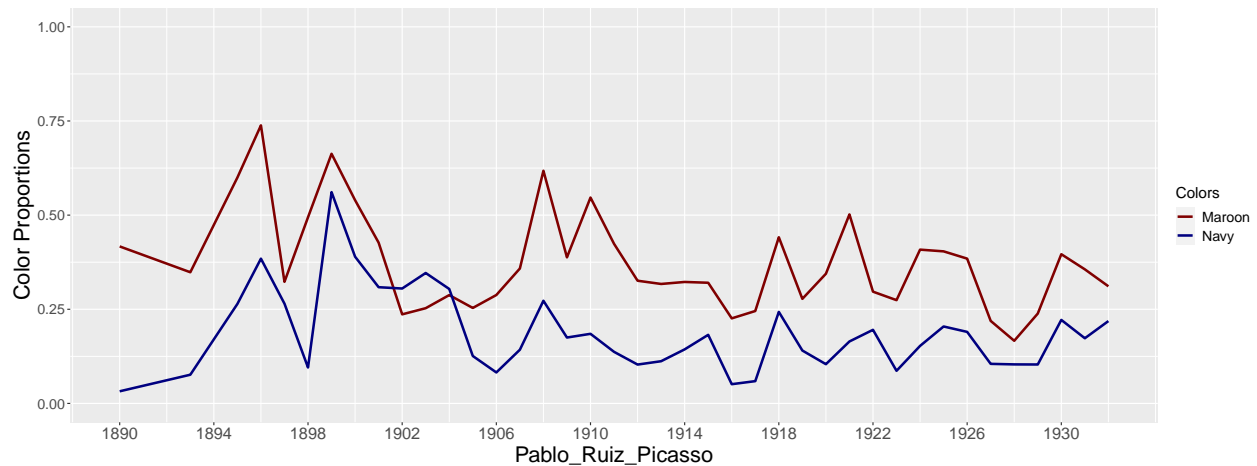


Based on the data collected for Salvador Dali(1913 - 1953),we can say from the above plot that the proportion of gray and olive was higher as compared to white and yellow throughout the years.



The above plot corresponds to the line diagram of average color proportions of eleven out of the sixteen basic colors considering all the paintings of Vincent Van Gogh from the years 1886 to 1890.

Remark: Study shows that in the paintings of Vincent Van Gogh, the proportion of blue and yellow was dominant. But the plot doesn't support the fact because the shades of blue and yellow used by Van Gogh were different from the yellow and blue we have considered.



The above plot corresponds to the line diagram of average color proportions of navy blue and brown colors considering all the paintings of Pablo Picasso from the years 1890 to 1930.

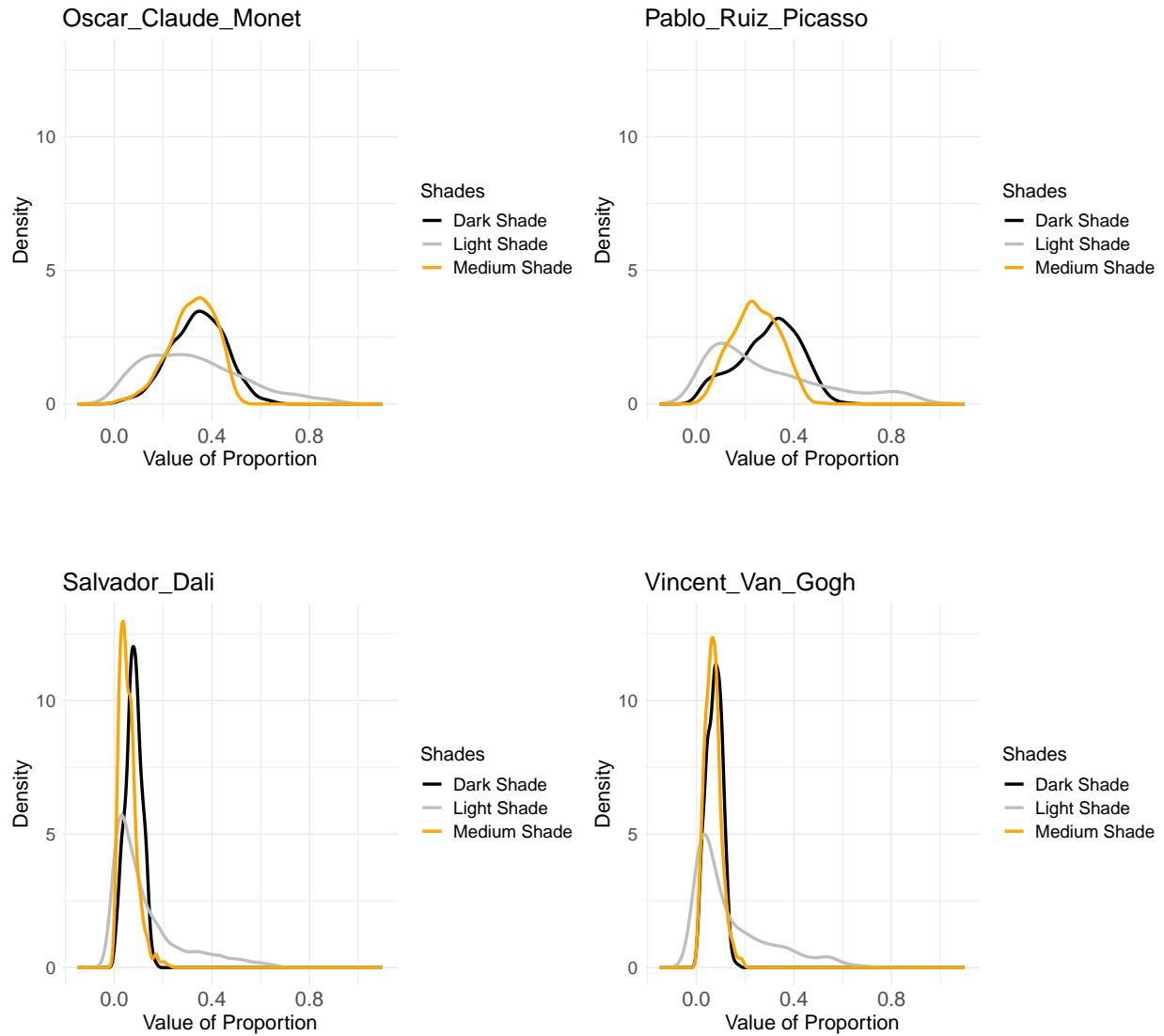
Interesting Fact:

Picasso's Blue Period: *"The Blue Period (Spanish: Período Azul) is a term used to define the works produced by Spanish painter Pablo Picasso between 1901 and 1904 when he painted essentially monochromatic paintings in shades of blue and blue-green, only occasionally warmed by other colors."* - taken from [https://en.wikipedia.org/wiki/Picasso%27s_Blue_Period#:~:text=The%20Blue%20Period%20\(Spanish%3A%20Per%C3%A9odo,occasionally%20warmed%20by%20other%20colors.](https://en.wikipedia.org/wiki/Picasso%27s_Blue_Period#:~:text=The%20Blue%20Period%20(Spanish%3A%20Per%C3%A9odo,occasionally%20warmed%20by%20other%20colors.)

The above plot also supports this fact, as we can see from the graph the values of blue color proportion are at its peak around 1900 to 1904 as compared to rest of the years.

Density Plot of Dark, Medium and Light Shades used in the Paintings Corresponding to Different Painters

First we have chosen 38 colors in dark shade category, 47 colors in medium category and 53 colors in light category. Then we computed the values of color proportions corresponding to all the colors in different shade categories for each paintings of the painters. After that we have taken mean of the color proportions corresponding to different color shades as the representative value of that corresponding shade category. The plots for different painters are given below,



Observing the above plots we can say that, on an average the proportion of dark and medium shades in Monet's paintings were around 0.2 to 0.5. But the variation in the proportion of light shade depicts that the proportion of light shades varies from painting to painting to comparatively large extent.

Similar observations can be made from the plot corresponding to Picasso.

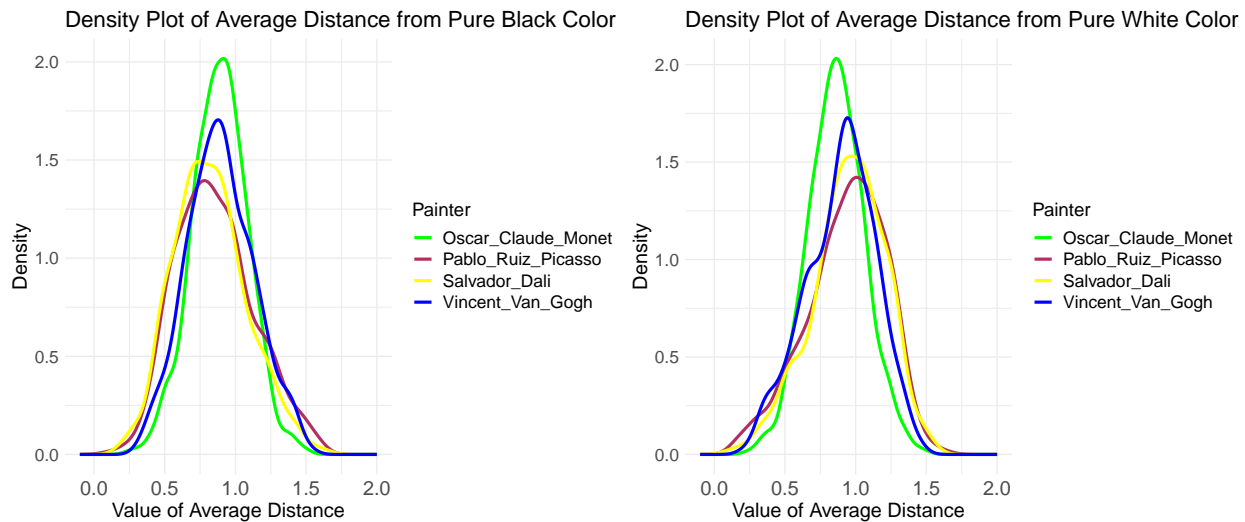
But from the plots of Dali and Van Gogh, it seems that they used very less proportion of all the three shades. This situation might have aroused due to insufficient choices of colors in different shade categories.

Density Plots of Average Distance of the Pixel Values From Pure White and Pure Black Color, Considering All the Paintings of the Painters

First we calculated the distance of all the pixels of the paintings from pure white and pure black color, so we get two different distance matrices(one for pure white and another for pure black color) corresponding different paintings of the painters. Then we take mean of each of the two distance matrices corresponding to different paintings, which we stored on two variables namely, “**dist.white**” and “**dist.black**”. Then we considered these two variables along with the first column of our original final data frame *final_data_grp3* to get another data frame named *white_black_df* which contains 2856 rows and 3 columns. A glimpse of this data set is given below,

```
##           Painter dist.white dist.black
## 1 Oscar_Claude_Monet 0.8437242 0.9039684
## 2 Oscar_Claude_Monet 1.4940344 0.2403546
## 3 Oscar_Claude_Monet 1.1506488 0.6168674
## 4 Oscar_Claude_Monet 1.2736792 0.5030720
## 5 Oscar_Claude_Monet 1.3559157 0.3893383
## 6 Oscar_Claude_Monet 0.9761023 0.7868317
```

On the basis of this new data frame we have made the following plots.



Here we can say that, the average distance of rgb pixels of the paintings from pure white and black colors corresponding to all the painters were clustered around 1.0 unit. But the peak of the density plots corresponding to Monet is higher as compared to others. This possibly means that Monet has larger number of paintings having almost similar proportion of white and black shades than the others.

A Brief Summary of R Shiny App

- The side panel of the R shiny consists of selections of painters, their corresponding years for which the data was collected, and a list of paintings for these selections. Then multiple group check boxes are provided for 16 basic colors. Lastly, one can choose painters for comparison.
- Corresponding to the user made selections, an image will load on the main panel with the following graphs; a density curve, a box plot of RGB color variation, color proportion plot, box plot & violin plot of median & quartile deviation for RGB color channels, line diagrams for average color proportions.
- User can choose from the various tabs we have provided to see different plots for different selections and make comparisons to get conclusions.

Final Conclusions of the Project

- Overlooking the density plot and box plot showing the variation in pixel values presented we can not tell all the possible colors used, but we can predict few of the colors that were predominantly used in the paintings.
- The color proportion plot we have presented shows the average proportion of 16 basic colors in a particular painting of a painter, and also in overall paintings of a painter from which we can say that all four of the painters used gray, silver and olive more as compared to rest of the 16 base colors. So color proportion plots can be used to see different color proportions in a painting of the painters. Overall color proportions can also be depicted by these plots. But as we are using only sixteen basic colors, there are some biases in the actual color proportions that have been discussed earlier.
- Observing the line diagrams of average color proportion of 16 basic colors throughout the years corresponding to each painter we can possibly say that the usage of gray was comparatively prominent. No particular color was used constantly throughout the years by any painter, if the usage of a color is more in an interval then its usage decreases or increases from that in the next few years. Thus, there was no consistency in the usage of colors.
- The extent to which different colors were used in a painting or in overall paintings of a given painter can be observed from box plots and violin plots presented but we can't name the exact shades of the colors.
- From the density plots of different categories of shades that is, dark, medium and light we can observe the average usage of these three shades considering all the paintings of each of the four painters.

References:

- <https://www.wikiart.org/>
- https://commons.wikimedia.org/wiki/Main_Page
- https://commons.wikimedia.org/wiki/Paintings_by_Claude_Monet
- <https://www.wikiart.org/en/pablo-picasso/all-works/text-list>
- <https://www.wikiart.org/en/salvador-dali/all-works/text-list>
- <https://www.wikiart.org/en/vincent-van-gogh/all-works/text-list>
- https://www.rapidtables.com/web/color/RGB_Color.html
- [https://en.wikipedia.org/wiki/Picasso%27s_Blue_Period#:~:text=The%20Blue%20Period%20\(Spanish%3A%20Per%C3%ADodo,occasionally%20warmed%20by%20other%20colors.](https://en.wikipedia.org/wiki/Picasso%27s_Blue_Period#:~:text=The%20Blue%20Period%20(Spanish%3A%20Per%C3%ADodo,occasionally%20warmed%20by%20other%20colors.)
- https://www.google.com/search?q=color+picker&rlz=1C1CHBF_enIN968IN968&oq=colo&aqs=chrome.1.69i57j69i59l3j0i67i131i433j0i433i512l2j46i67i175i199j46i131i433i512j0i433.1917j0j15&sourceid=chrome&ie=UTF-8