

Stat 480 - Homework #8

Vahid Azizi

3/12/2020

The Joy of Painting with Bob Ross



Submission Details

Due date: the homework is due before class on Thursday.

Submission process: submit both the R Markdown file and the corresponding html file on canvas. Please submit both the .Rmd and the .html files separately and do not zip the two files together.

Questions

1. Download the RMarkdown file with these homework instructions to use as a template for your work. Make sure to replace "Your Name" in the YAML with your name.

```
# Load necessary Libraries
library(dplyr)
library(tidy)
library(ggplot2)
library(forcats)
library(readr)
```

2. The data this week comes from fivethirtyeight (<https://github.com/fivethirtyeight/data>). The data set includes information on the 403 episodes of "The Joy of Painting". The accompanying article is published here (<https://fivethirtyeight.com/features/a-statistical-analysis-of-the-work-of-bob-ross/>).

```
# read in the data
bob_ross <- read_csv('https://raw.githubusercontent.com/Stat480-at-ISU/Stat480-at-ISU.github.io/master/homework/data/bob-ross.csv')
```

3. Variables `alizarin_crimson` through `burnt_umber` correspond to the binary presence (0 or 1) of that color in the painting. Gather all of these variables and create a long form of the data, introducing two new variables called `color` and `presence`. Save the result in a data frame called `bob_ross_colors`.

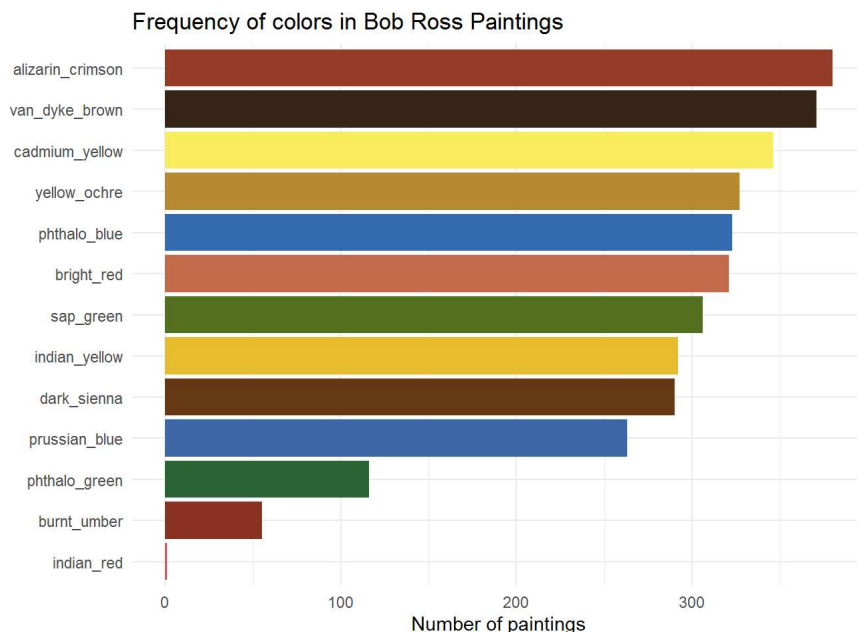
```
columnnames<-t(t(colnames(bob_ross)))
sel_columns<-columnnames[53:65]
bob_ross_colors<-pivot_longer(bob_ross,cols=sel_columns,names_to = "color",values_to = "presence")
head(bob_ross_colors)
```

```
## # A tibble: 6 x 54
##   season episode title aurora_borealis barn beach boat bridge building bushes
##   <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1     1 1 A Wa~           0     0     0     0     0     0     1
## 2     1     1 1 A Wa~           0     0     0     0     0     0     1
## 3     1     1 1 A Wa~           0     0     0     0     0     0     1
## 4     1     1 1 A Wa~           0     0     0     0     0     0     1
## 5     1     1 1 A Wa~           0     0     0     0     0     0     1
## 6     1     1 1 A Wa~           0     0     0     0     0     0     1
## # ... with 44 more variables: cabin <dbl>, cactus <dbl>, cirrus <dbl>,
## # cliff <dbl>, clouds <dbl>, conifer <dbl>, cumulus <dbl>, deciduous <dbl>,
## # dock <dbl>, farm <dbl>, fence <dbl>, fire <dbl>, flowers <dbl>, fog <dbl>,
## # grass <dbl>, guest <dbl>, hills <dbl>, lake <dbl>, lakes <dbl>,
## # lighthouse <dbl>, mill <dbl>, moon <dbl>, mountain <dbl>, mountains <dbl>,
## # night <dbl>, ocean <dbl>, palm_trees <dbl>, path <dbl>, person <dbl>,
## # portrait <dbl>, river <dbl>, rocks <dbl>, snow <dbl>, snowy_mountain <dbl>,
## # structure <dbl>, sun <dbl>, tree <dbl>, trees <dbl>, waterfall <dbl>,
## # waves <dbl>, windmill <dbl>, winter <dbl>, color <chr>, presence <dbl>
```

4. Does Bob Ross have a favorite color to paint with? Use the data `bob_ross_colors` as your starting point and for each color calculate the number of times that color was used throughout the series. After using this number to reorder the levels of the variable `color`, create a bar chart using the code below as your starting point and add in the necessary aesthetic mappings within `ggplot(aes())`. Describe and summarize the chart.

```
question4_data %>%
  ggplot(aes( )) +
  geom_bar(show.legend = FALSE) +
  coord_flip() +
  theme_minimal() +
  labs(y = "Number of paintings",
       x = "",
       title = "Frequency of colors in Bob Ross Paintings") +
  scale_fill_manual(values = c("#CD5C5C", "#8A3324", "#2C6436", "#3C67A7", "#643914", "#E7BD2F", "#546F1F", "#C36A4A", "#346
BB1", "#B58A30", "#F8ED5F", "#372518", "#973B29"))
```

```
bob_ross_colors %>% filter(presence==1) %>% group_by(color) %>%
  tally %>% arrange(desc(n)) %>%
  ggplot(aes(y=n,x=fct_reorder(color,n),fill = fct_reorder(color,n))) +
  geom_bar(stat = "identity",show.legend = FALSE) +
  coord_flip() +
  theme_minimal() +
  labs(y = "Number of paintings",
       x = "",
       title = "Frequency of colors in Bob Ross Paintings") +
  scale_fill_manual(values = c("#CD5C5C", "#8A3324", "#2C6436", "#3C67A7", "#643914", "#E7BD2F", "#546F1F", "#C36A4A", "#346
BB1", "#B58A30", "#F8ED5F", "#372518", "#973B29"))
```



This bar chart aims to show the most favorite colors of Bob Ross in his paintings. As can be seen in this bar chart, `alizarin_crimson`, `van_dyke_brown`, and `cadmium_yellow` are top3 with more than 300 times presence in his paintings. Also, `indian_red` is the least favorite one with just 1 presence in all seasons.

5. For this question use the original data again. Variables `aurora_borealis` through `winter` correspond to the binary presence (0 or 1) of that element in the painting. Use `pivot_longer()` as shown in class to transform the data into a tidier format with new variables `element` and `presence`. Save the result in a data frame called `bob_ross_elements`.

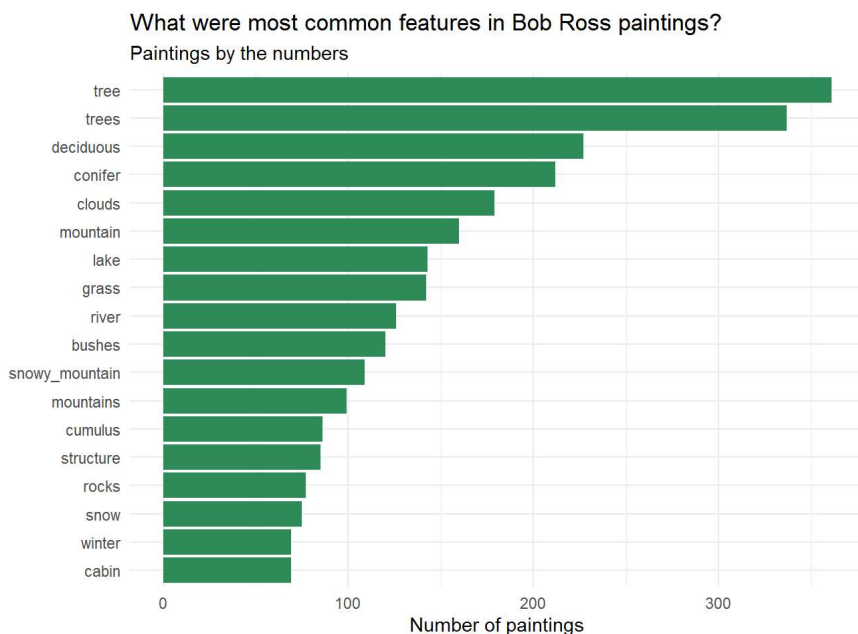
```
element_columns<-columnnames[4:52]
bob_ross_elements<-pivot_longer(bob_ross,cols=element_columns,names_to = "element",values_to = "presence")
head(bob_ross_elements)
```

```
## # A tibble: 6 x 18
##   season episode title alizarin_crimson bright_red cadmium_yellow dark_sienna
##   <dbl>   <dbl> <chr>          <dbl>      <dbl>      <dbl>      <dbl>
## 1     1     1 1 A Wa~           1         1         1         0
## 2     1     1 1 A Wa~           1         1         1         0
## 3     1     1 1 A Wa~           1         1         1         0
## 4     1     1 1 A Wa~           1         1         1         0
## 5     1     1 1 A Wa~           1         1         1         0
## 6     1     1 1 A Wa~           1         1         1         0
## # ... with 11 more variables: indian_yellow <dbl>, phthalo_blue <dbl>,
## #   phthalo_green <dbl>, prussian_blue <dbl>, sap_green <dbl>,
## #   van_dyke_brown <dbl>, yellow_ochre <dbl>, indian_red <dbl>,
## #   burnt_umber <dbl>, element <chr>, presence <dbl>
```

6. What are the most frequent elements in his paintings? Use the data `bob_ross_elements` as your starting point and for each element calculate the number of times that element was included. Then use this number to reorder the levels of `element`. Exclude elements that were featured in fewer than 50 paintings and create a bar chart. Use the code below as your starting point and add in the necessary aesthetic mappings within `ggplot(aes())`. Describe and summarize the chart.

```
question6_data %>%
  ggplot(aes( )) +
  geom_bar(fill = "seagreen") +
  coord_flip() +
  theme_minimal() +
  labs(y = "Number of paintings",
       x = "",
       title = "What were most common features in Bob Ross paintings?",
       subtitle = "Paintings by the numbers")
```

```
bob_ross_elements %>% filter(presence==1) %>% group_by(element) %>%
  tally %>% arrange(desc(n)) %>% filter(n>50) %>%
  ggplot(aes(x=fct_reorder(element,n),y=n)) +
  geom_bar(stat = "identity",fill = "seagreen") +
  coord_flip() +
  theme_minimal() +
  labs(y = "Number of paintings",
       x = "",
       title = "What were most common features in Bob Ross paintings?",
       subtitle = "Paintings by the numbers")
```



This bar chart seeks to show the

most favorite elements Bob Ross used in his paintings. According to this figure, tree, trees are the most favorite elements with more than 350 times usage. Also, cabin is the least frequent item in his paintings.

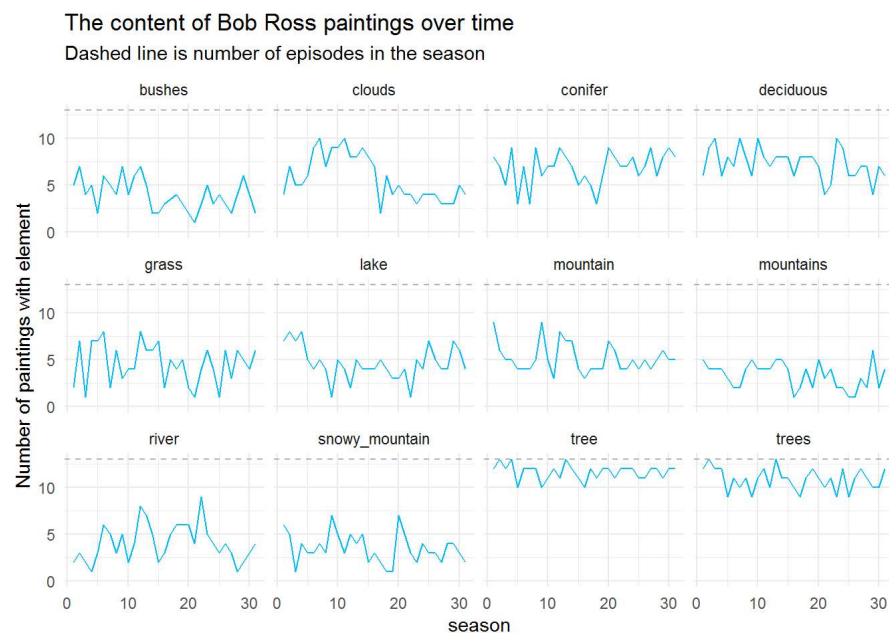
7. Did the content of the paintings change over time? We will attempt to answer this question by looking at elements that appeared in more than 90 paintings and their trends over the seasons. Use the data `bob_ross_elements` as your starting point and for each season and element, calculate the number of times an element was included. Exclude elements that were included in less than 90 paintings total. Create a line plot showing number of times an element was included for each season with season on the x-axis and facet by element. Use the code below as your starting point and add in the necessary aesthetic mappings within `ggplot(aes())` and add in the faceting. Describe and summarize the chart.

```
question7_data %>% ggplot(aes( )) + geom_line(color = "deepskyblue") + # add faceting here labs(y = "Number of paintings with element",
```

```
most_repeated_elemsnts<-bob_ross_elements %>% filter(presence==1)%>%group_by(element)%>%
  tally %>% filter(n>90)

favorite_elements<-unique(most_repeated_elemsnts$element)

bob_ross_elements %>% filter(presence==1,element %in% favorite_elements) %>%group_by(season,element)%>% tally %>%
  ggplot(aes(x=season,y=n )) +
  geom_line(color = "deepskyblue") +
  facet_wrap(~element)+
  labs(y = "Number of paintings with element",
       title = "The content of Bob Ross paintings over time",
       subtitle = "Dashed line is number of episodes in the season") +
  geom_hline(yintercept = 13, lty = 2, color = "grey70") +
  theme_minimal() +
  expand_limits(y = 0)
```



These plots aim to show the trend of appearance for different elements in Bob Ross's paintings during different seasons. As can be seen, for most of the elements, the line plot fluctuates and no specific trend can be recognized. But, in some of them there are some trends. For example, in clouds plot, it can be seen that in the last seasons the number of times clouds were used had been decreased. Lake also starts with high number of appearance but in the later seasons its frequency decreases.