ETC3250 - Project

Souhaib Ben Taieb, Di Cook, Rob Hyndman

The project for this course will be to analyse paintings by Bob Ross. This was the subject of the 538 post.

We have taken the painting images from the sales site, read the images into R, resized them all to be 20 by 20 pixels. Each painting has been classified into one of 8 classes. This is the data that you will work with.

It is provided in wide and long form. Long form is good for making pictures of the original painting, and the wide form is what you will need to use for the classification. In wide form, each row corresponds to one painting, and the rgb color values at each pixel are in each column. With a 20x20 image, this leads to 400x3=1200 columns.

There are 241 paintings in the full data set. We have given you a training set of 178 paintings. Your job is to build the best classifier that you can using this training data and use it to predict the data that we have not labelled. We have the classes for this test data, and can provide you with the error for your predictions.

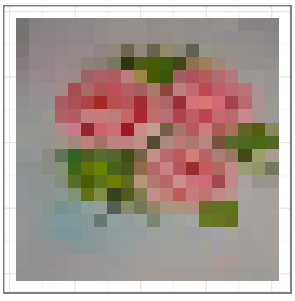
Here is a quick look at the data, and making plots.

```
## [1]  178 1203
```

```
##
##        cold        dusk     flowers impressions        oval       scene
##          23          30          22          17           5          28
##       trees       water
##          18          35
```

```
## [1] 71200     9
```

```
## [1] "scene"   "water"   "flowers"
```



These are the original paintings for these three data points.



The labels for the paintings have been automatically created from the paintings names. It is possible that a painting might be better labelled as a different class.

Deadlines:

Grading:

Uploading predictions: