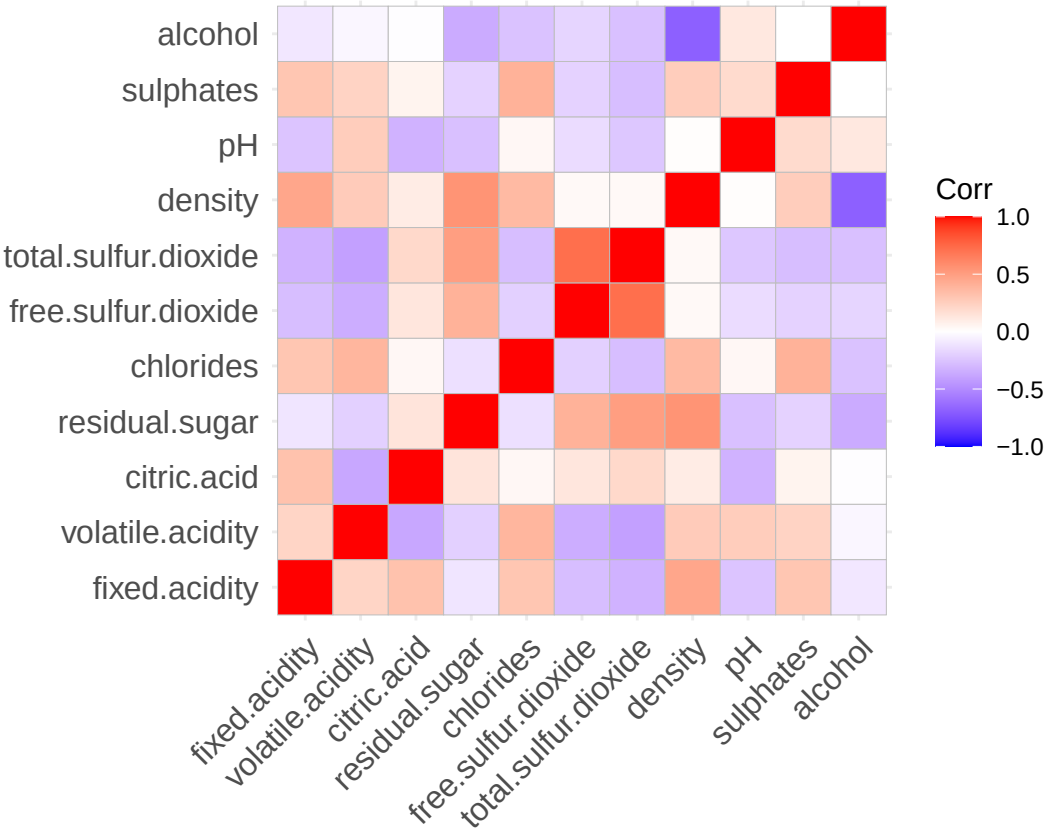
Clustering and Dimensionality Reduction

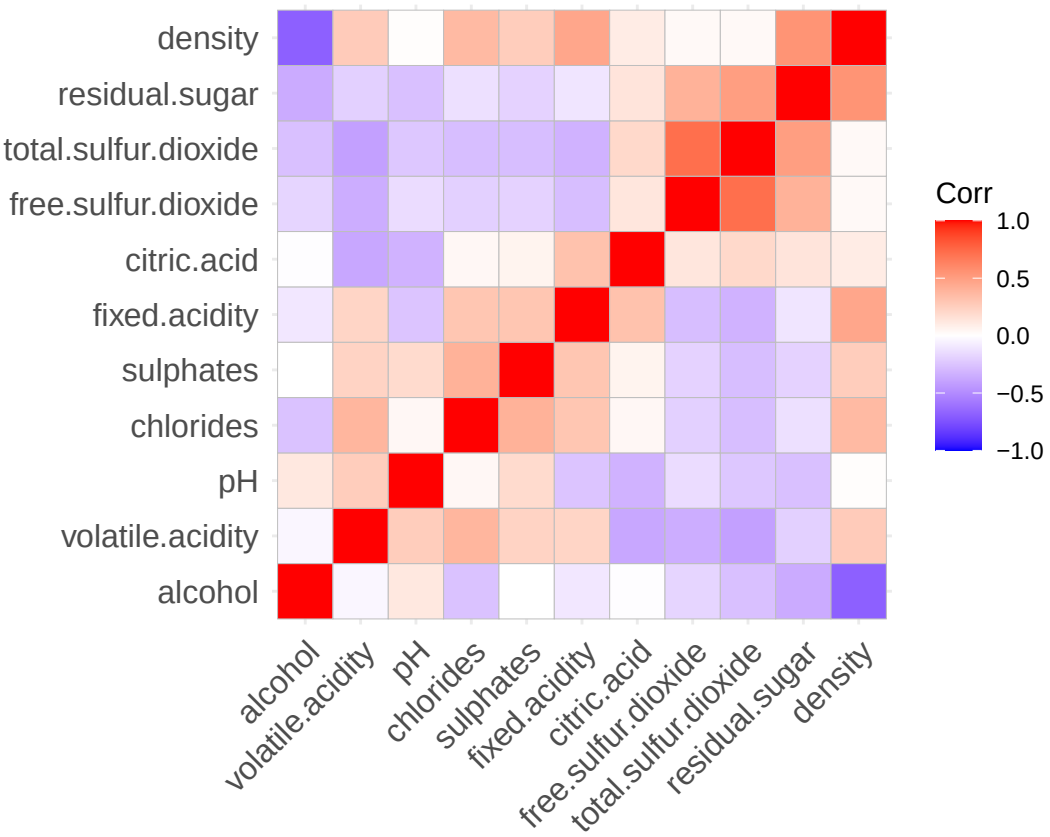
2023-08-11

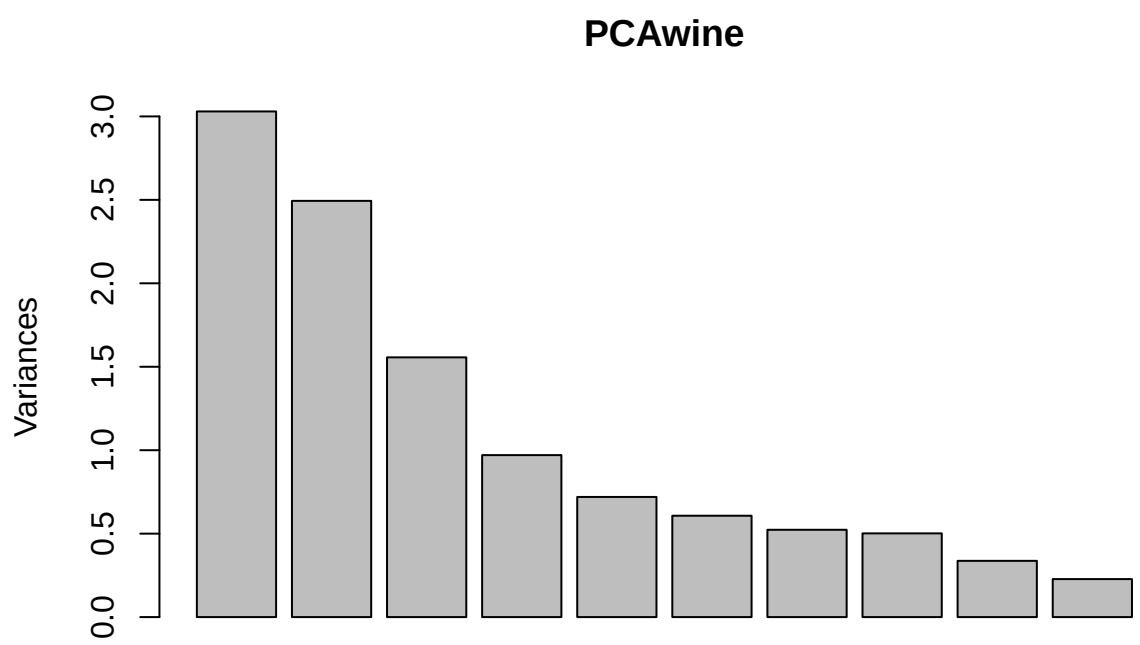
```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
## Loading required package: lattice
##
##
## Attaching package: 'caret'
##
##
## The following object is masked from 'package:purrr':
##
##     lift
##
##
## Loading required package: grid
##
## Loading required package: modeltools
##
## Loading required package: stats4
##
##
## Attaching package: 'foreach'
##
##
## The following objects are masked from 'package:purrr':
##
##     accumulate, when
```

```
##                       fixed.acidity volatile.acidity citric.acid residual.sugar
## fixed.acidity           1.00000000       0.21900826  0.32443573     -0.1119813
## volatile.acidity        0.21900826       1.00000000 -0.37798132     -0.1960112
## citric.acid             0.32443573      -0.37798132  1.00000000      0.1424512
## residual.sugar         -0.11198128      -0.19601117  0.14245123      1.0000000
## chlorides               0.29819477       0.37712428  0.03899801     -0.1289405
## free.sulfur.dioxide    -0.28273543      -0.35255731  0.13312581      0.4028706
## total.sulfur.dioxide   -0.32905390      -0.41447619  0.19524198      0.4954816
## density                 0.45890998       0.27129565  0.09615393      0.5525170
## pH                     -0.25270047       0.26145440 -0.32980819     -0.2673198
```

```
## sulphates               0.29956774       0.22598368  0.05619730     -0.1859274
## alcohol                -0.09545152      -0.03764039 -0.01049349     -0.3594148
##                         chlorides free.sulfur.dioxide total.sulfur.dioxide
## fixed.acidity          0.29819477         -0.28273543          -0.32905390
## volatile.acidity       0.37712428         -0.35255731          -0.41447619
## citric.acid            0.03899801          0.13312581           0.19524198
## residual.sugar        -0.12894050          0.40287064           0.49548159
## chlorides              1.00000000         -0.19504479          -0.27963045
## free.sulfur.dioxide   -0.19504479          1.00000000           0.72093408
## total.sulfur.dioxide  -0.27963045          0.72093408           1.00000000
## density                0.36261466          0.02571684           0.03239451
## pH                     0.04470798         -0.14585390          -0.23841310
## sulphates              0.39559331         -0.18845725          -0.27572682
## alcohol               -0.25691558         -0.17983843          -0.26573964
##                           density           pH    sulphates      alcohol
## fixed.acidity          0.45890998 -0.25270047  0.299567744 -0.095451523
## volatile.acidity       0.27129565  0.26145440  0.225983680 -0.037640386
## citric.acid            0.09615393 -0.32980819  0.056197300 -0.010493492
## residual.sugar         0.55251695 -0.26731984 -0.185927405 -0.359414771
## chlorides              0.36261466  0.04470798  0.395593307 -0.256915580
## free.sulfur.dioxide    0.02571684 -0.14585390 -0.188457249 -0.179838435
## total.sulfur.dioxide   0.03239451 -0.23841310 -0.275726820 -0.265739639
## density                1.00000000  0.01168608  0.259478495 -0.686745422
## pH                     0.01168608  1.00000000  0.192123407  0.121248467
## sulphates              0.25947850  0.19212341  1.000000000 -0.003029195
## alcohol               -0.68674542  0.12124847 -0.003029195  1.000000000
```




density
residual.sugar
total.sulfur.dioxide
free.sulfur.dioxide
citric.acid
fixed.acidity
sulphates
chlorides
pH
volatile.acidity
alcohol
alcohol
volatile.acidity
pH
chlorides
sulphates
fixed.acidity
citric.acid
free.sulfur.dioxide
total.sulfur.dioxide
residual.sugar
density
Corr
1.0
0.5
0.0
−0.5
−1.0



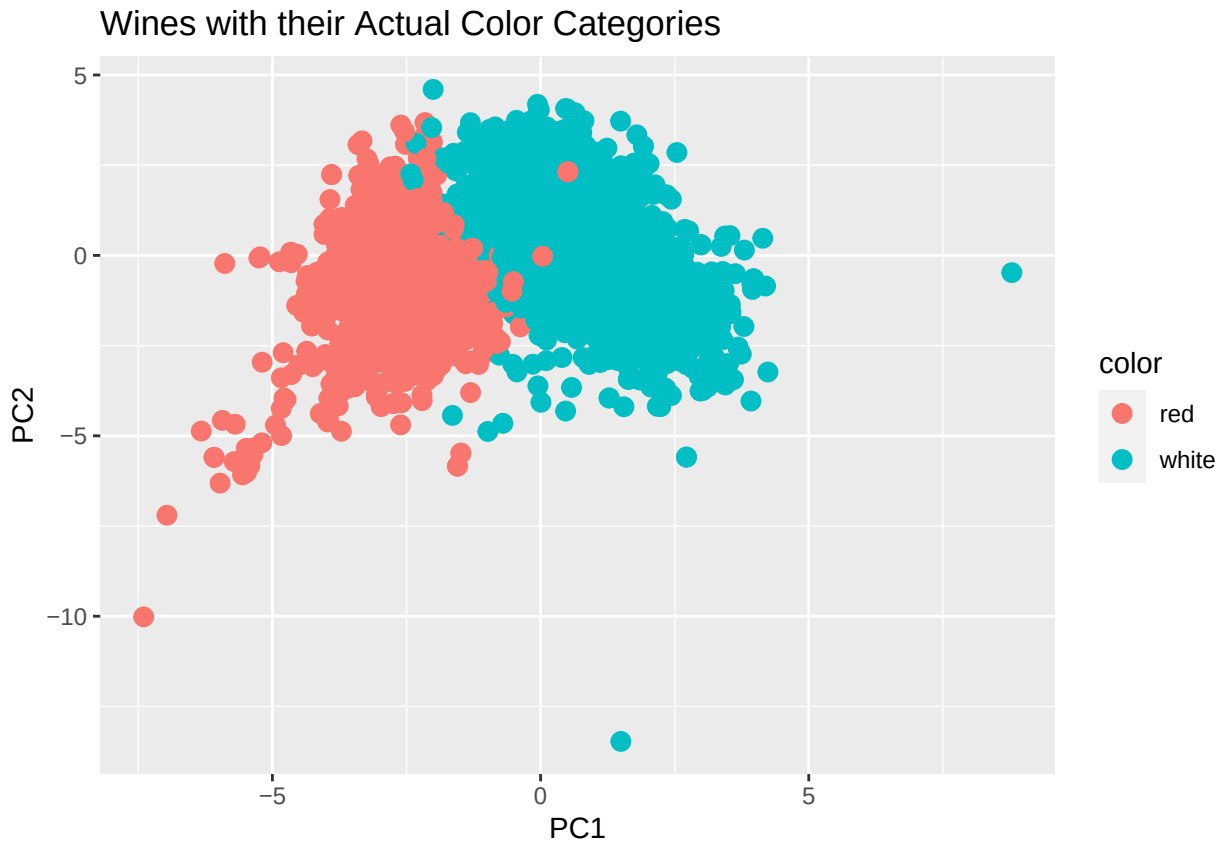

```
## Importance of first k=2 (out of 11) components:
##                            PC1    PC2
## Standard deviation     1.7407 1.5792
## Proportion of Variance 0.2754 0.2267
## Cumulative Proportion  0.2754 0.5021
```

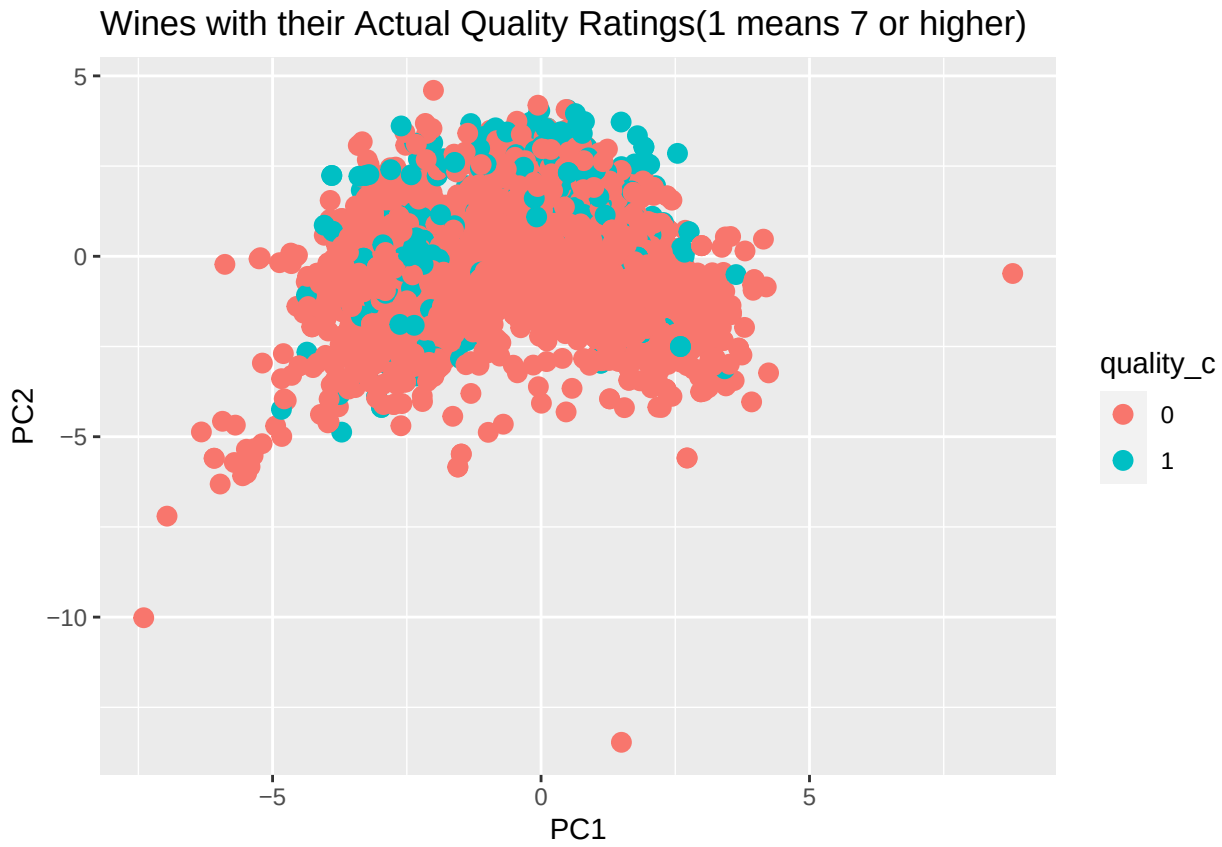
```
##                        PC1   PC2
## fixed.acidity        -0.24 -0.34
## volatile.acidity     -0.38 -0.12
## citric.acid           0.15 -0.18
## residual.sugar        0.35 -0.33
## chlorides            -0.29 -0.32
## free.sulfur.dioxide   0.43 -0.07
## total.sulfur.dioxide  0.49 -0.09
## density              -0.04 -0.58
## pH                   -0.22  0.16
## sulphates            -0.29 -0.19
## alcohol              -0.11  0.47
```

```
##                Question         PC1
## 1  total.sulfur.dioxide  0.48741806
## 2   free.sulfur.dioxide  0.43091401
## 3        residual.sugar  0.34591993
## 4           citric.acid  0.15238844
## 5               density -0.04493664
```

```
## 6               alcohol -0.10643712
## 7                    pH -0.21868644
## 8         fixed.acidity -0.23879890
## 9             chlorides -0.29011259
## 10            sulphates -0.29413517
## 11     volatile.acidity -0.38075750
```

```
##                Question         PC2
## 1               alcohol  0.46505769
## 2                    pH  0.15586900
## 3   free.sulfur.dioxide -0.07193260
## 4  total.sulfur.dioxide -0.08726628
## 5      volatile.acidity -0.11754972
## 6           citric.acid -0.18329940
## 7             sulphates -0.19171577
## 8             chlorides -0.31525799
## 9       residual.sugar -0.32991418
## 10        fixed.acidity -0.33635454
## 11              density -0.58403734
```





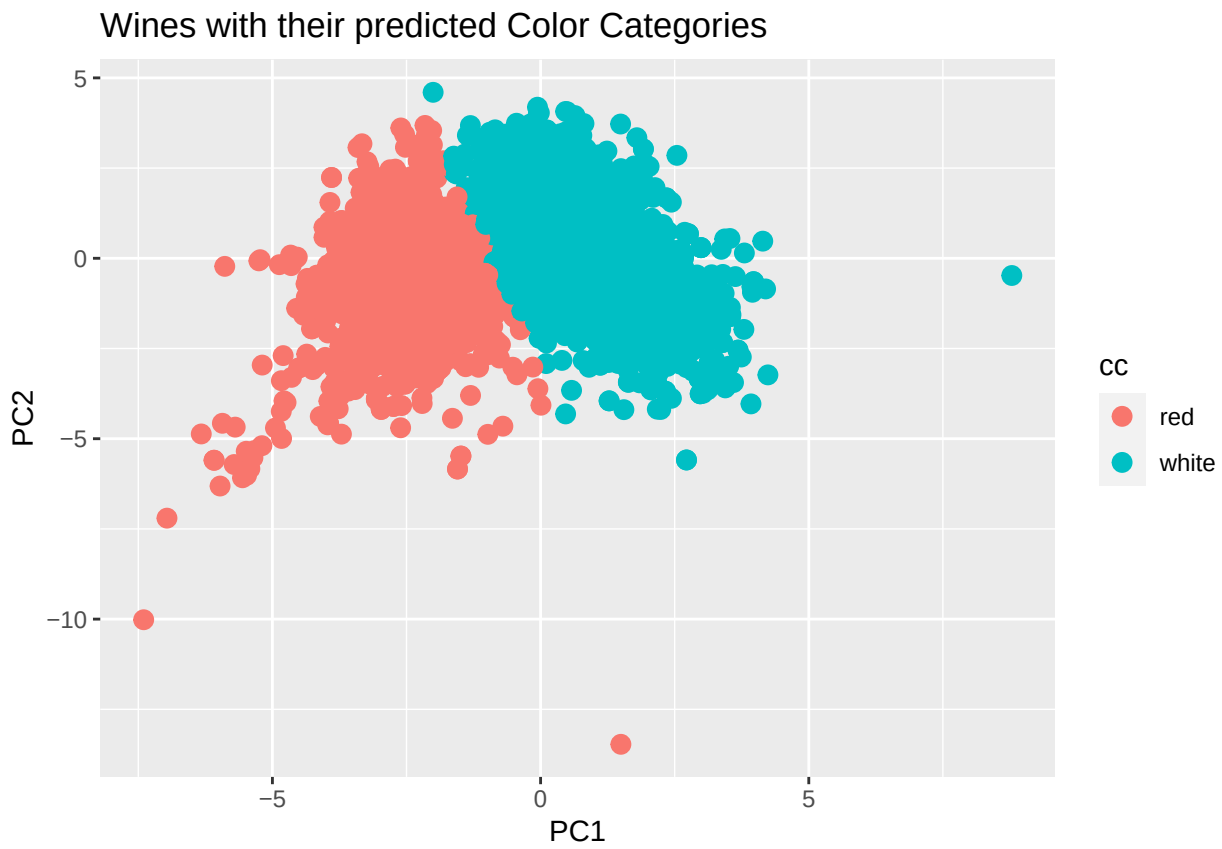

```
## [1] 35881.92
```

```
## [1] 15122.33
```

```
## [1] 15912.783  4846.805
```

```
## [1] 4841 1656
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.255   2.000   2.000
```




```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  red white
##      red   1572    84
##      white   27  4814
##
##                Accuracy : 0.9829
##                  95% CI : (0.9795, 0.9859)
##     No Information Rate : 0.7539
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9545
##
##  Mcnemar's Test P-Value : 1.065e-07
##
##             Sensitivity : 0.9831
##             Specificity : 0.9829
##          Pos Pred Value : 0.9493
##          Neg Pred Value : 0.9944
##              Prevalence : 0.2461
##          Detection Rate : 0.2420
##    Detection Prevalence : 0.2549
##       Balanced Accuracy : 0.9830
##
##        'Positive' Class : red
```

##