

Data Analytics: Lab Heatmaps, Dimension Reduction, PCA

Brendan Donnelly

October 2, 2020

```
wine_data <- read.table("wine.data", sep = ",")

colnames(wine_data) <- c("Cvs", "Alcohol", "Malic_Acid", "Ash", "Alcalinity_of_Ash",
  "Magnesium", "Total_Phenols", "Flavanoids", "NonFlavanoidPhenols",
  "Proanthocyanins", "Color_Intenstity", "Hue", "OD280/OD315_of_Diluted_Wine",
  "Proline")

head(wine_data)
```

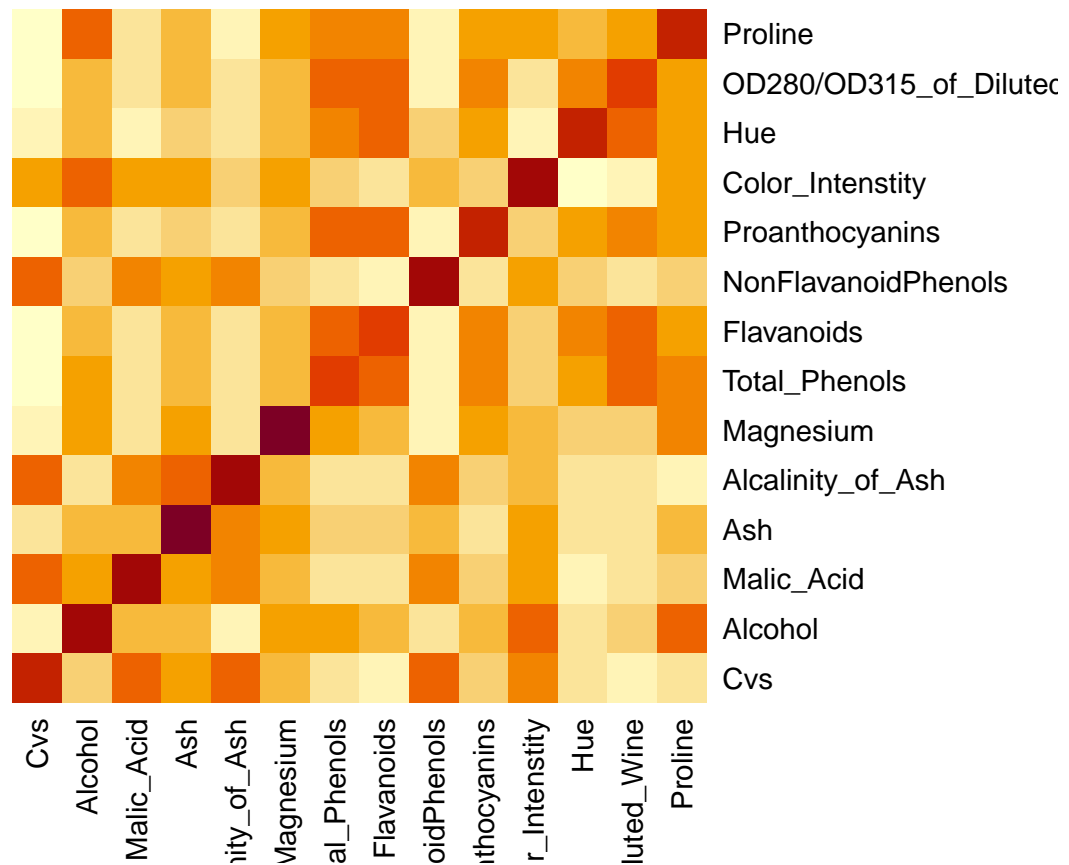
##	Cvs	Alcohol	Malic_Acid	Ash	Alcalinity_of_Ash	Magnesium	Total_Phenols
## 1	1	14.23	1.71	2.43	15.6	127	2.80
## 2	1	13.20	1.78	2.14	11.2	100	2.65
## 3	1	13.16	2.36	2.67	18.6	101	2.80
## 4	1	14.37	1.95	2.50	16.8	113	3.85
## 5	1	13.24	2.59	2.87	21.0	118	2.80
## 6	1	14.20	1.76	2.45	15.2	112	3.27
##	Flavanoids	NonFlavanoidPhenols	Proanthocyanins	Color_Intenstity	Hue		
## 1	3.06	0.28	2.29	5.64	1.04		
## 2	2.76	0.26	1.28	4.38	1.05		
## 3	3.24	0.30	2.81	5.68	1.03		
## 4	3.49	0.24	2.18	7.80	0.86		
## 5	2.69	0.39	1.82	4.32	1.04		
## 6	3.39	0.34	1.97	6.75	1.05		
##	OD280/OD315_of_Diluted_Wine	Proline					
## 1	3.92	1065					
## 2	3.40	1050					
## 3	3.17	1185					
## 4	3.45	1480					
## 5	2.93	735					
## 6	2.85	1450					

```
tail(wine_data)
```

##	Cvs	Alcohol	Malic_Acid	Ash	Alcalinity_of_Ash	Magnesium	Total_Phenols
## 173	3	14.16	2.51	2.48	20.0	91	1.68
## 174	3	13.71	5.65	2.45	20.5	95	1.68
## 175	3	13.40	3.91	2.48	23.0	102	1.80
## 176	3	13.27	4.28	2.26	20.0	120	1.59
## 177	3	13.17	2.59	2.37	20.0	120	1.65
## 178	3	14.13	4.10	2.74	24.5	96	2.05
##	Flavanoids	NonFlavanoidPhenols	Proanthocyanins	Color_Intenstity	Hue		
## 173	0.70	0.44	1.24	9.7	0.62		
## 174	0.61	0.52	1.06	7.7	0.64		
## 175	0.75	0.43	1.41	7.3	0.70		

```
## 176      0.69      0.43      1.35      10.2 0.59
## 177      0.68      0.53      1.46      9.3 0.60
## 178      0.76      0.56      1.35      9.2 0.61
##      OD280/OD315_of_Diluted_Wine Proline
## 173      1.71      660
## 174      1.74      740
## 175      1.56      750
## 176      1.56      835
## 177      1.62      840
## 178      1.60      560
```

```
heatmap(cor(wine_data),Rowv=NA,Colv=NA)
```



```
cultivar_classes <- factor(wine_data$Cvs)
```

```
wine_data_pca <- prcomp(scale(wine_data[, -1]))
summary(wine_data_pca)
```

```
## Importance of components:
##              PC1      PC2      PC3      PC4      PC5      PC6      PC7
## Standard deviation  2.169 1.5802 1.2025 0.95863 0.92370 0.80103 0.74231
## Proportion of Variance 0.362 0.1921 0.1112 0.07069 0.06563 0.04936 0.04239
## Cumulative Proportion 0.362 0.5541 0.6653 0.73599 0.80162 0.85098 0.89337
##              PC8      PC9      PC10      PC11      PC12      PC13
## Standard deviation  0.59034 0.53748 0.5009 0.47517 0.41082 0.32152
## Proportion of Variance 0.02681 0.02222 0.0193 0.01737 0.01298 0.00795
## Cumulative Proportion 0.92018 0.94240 0.9617 0.97907 0.99205 1.00000
```