

Logistic Regression

Karol Orozco & Charles Hanks

text feature engineering:

```
wine_words <- function(df, j = 1000, stem=F){
  library(tidytext)
  library(SnowballC)
  data(stop_words)
  words <- df %>%
    unnest_tokens(word, description) %>%
    anti_join(stop_words) %>% # get rid of stop words
    filter(!(word %in% c("wine", "pinot", "vineyard", "price", "points")))

  if(stem){
    words <- words %>%
      mutate(word = wordStem(word))
  }

  words <- words %>%
    count(id, word) %>%
    group_by(id) %>%
    mutate(exists = (n>0)) %>%
    ungroup %>%
    group_by(word) %>%
    mutate(total = sum(n)) %>%
    filter(total > j) %>%
    pivot_wider(id_cols = id, names_from = word, values_from = exists, values_fill = list(
    right_join(select(df, id, province)) %>%
    mutate(across(-province, ~replace_na(.x, F)))
  )
  wino <- wine_words(wine, j=400, stem=F)
```

Joining with `by = join_by(word)`

Joining with `by = join_by(id)`

bringing back numerical features from original dataset to wino:

```
wino = wino %>% left_join(select(wine, id, price, points, year), by = "id")
```

Numerical feature engineering:

```
#center and scale points:
wino = wino %>% select(points) %>% preProcess(method = c("center", "scale")) %>% predict(w

#year as factor, logprice:
wino = wino %>% mutate(year_f = as.factor(year),
                        lprice = log(price))

#binning year and price:
wino = wino %>%
  mutate(price_f = case_when(
    price < 16 ~ "low",
    price >= 16 & price < 41 ~ "med",
    price >= 41 ~ "high"
  ),
  year_f = case_when(
    year < 2005 ~ "old",
    year >= 2005 & year < 2011 ~ "recent",
    year >= 2011 ~ "current"
  ))
wino = wino %>% dplyr::select(-price)
#difference of wine's lprice from total average lprice
wino = wino %>% mutate(diff_from_avg_lprice = mean(lprice) - lprice)
wino = wino %>% mutate(cost_per_point = lprice/points)
wino = wino %>% select(-id)
wino = wino %>% select(-diff_from_avg_lprice)

head(wino)
```

A tibble: 6 x 82

	bottling	earthy	herbal	berry	chocolate	drink	herb	oak	tart	aromas	bodied
	<lgl>	<lgl>	<lgl>	<lgl>	<lgl>	<lgl>	<lgl>	<lgl>	<lgl>	<lgl>	<lgl>
1	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
2	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE
3	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE

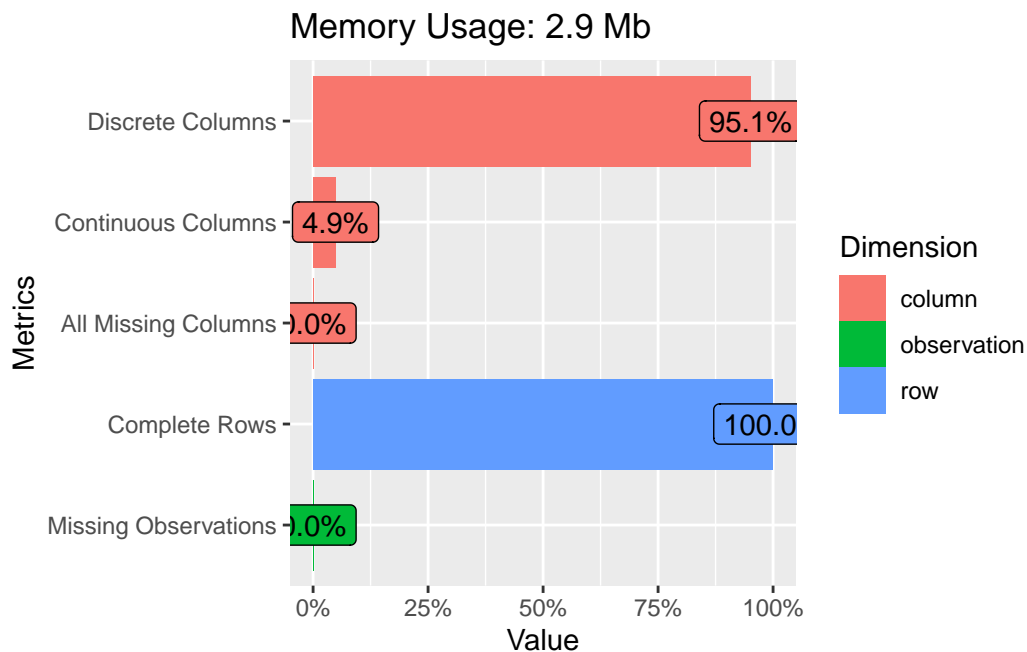
```

4 FALSE    FALSE TRUE    FALSE FALSE    FALSE TRUE  FALSE FALSE FALSE  FALSE
5 FALSE    TRUE  FALSE TRUE    FALSE    FALSE FALSE FALSE FALSE FALSE  FALSE
6 FALSE    FALSE FALSE FALSE FALSE    TRUE  FALSE FALSE FALSE FALSE  FALSE
# ... with 71 more variables: earth <lgl>, forest <lgl>, offers <lgl>,
#   raspberry <lgl>, smooth <lgl>, spice <lgl>, texture <lgl>, finish <lgl>,
#   flavor <lgl>, fruit <lgl>, notes <lgl>, sweet <lgl>, touch <lgl>,
#   flavors <lgl>, tannins <lgl>, fruity <lgl>, strawberry <lgl>,
#   cranberry <lgl>, dark <lgl>, palate <lgl>, acidity <lgl>, black <lgl>,
#   cherry <lgl>, cola <lgl>, dried <lgl>, nose <lgl>, soft <lgl>, juicy <lgl>,
#   ripe <lgl>, light <lgl>, spicy <lgl>, red <lgl>, age <lgl>, bit <lgl>, ...

```

```
library("DataExplorer")
```

```
plot_intro(wino)
```



Split data

```

set.seed(100)
wine_index <- createDataPartition(wino$province, p = 0.80, list = FALSE)

train <- wino[ wine_index, ]

```

```
test <- wino[-wine_index, ]
```

```
table(train$province)
```

Burgundy	California	Casablanca_Valley	Marlborough
955	3168	105	184
New_York	Oregon		
105	2190		

```
nrow(train)
```

```
[1] 6707
```

```
nrow(test)
```

```
[1] 1673
```

Fit Model

```
# Fit the mode
```

```
control <- trainControl(method="cv",  
                        number=10,  
                        savePredictions="all",  
                        classProbs=TRUE)
```

```
model <- nnet::multinom(province ~.,  
                        data = train,  
                        trControl=control)
```

```
# weights: 510 (420 variable)  
initial value 12017.330760  
iter 10 value 8583.350102  
iter 20 value 3447.156328  
iter 30 value 2928.883427  
iter 40 value 2399.319643  
iter 50 value 2184.899194
```

```

iter 60 value 2062.258244
iter 70 value 1942.942586
iter 80 value 1897.121134
iter 90 value 1846.706928
iter 100 value 1823.360759
final value 1823.360759
stopped after 100 iterations

```

```
print(model)
```

Call:

```
nnet::multinom(formula = province ~ ., data = train, trControl = control)
```

Coefficients:

	(Intercept)	bottlingTRUE	earthyTRUE	herbalTRUE	berryTRUE
California	-2.04158110	2.3326692	2.33442850	2.6184997	-1.48538107
Casablanca_Valley	-0.06739038	0.4766752	1.61014621	4.4720869	1.16357544
Marlborough	0.27853449	2.1998386	-0.05894852	2.8450814	-2.26217716
New_York	-0.05014946	-2.1480213	1.15189817	0.2317209	0.02374801
Oregon	1.42782315	1.3700733	0.98981926	2.6651440	-0.54422312
	chocolateTRUE	drinkTRUE	herbTRUE	oakTRUE	tartTRUE
California	-0.166739656	-3.7928243	3.533443	3.913317	3.450648
Casablanca_Valley	3.226693701	-1.0833661	1.330802	4.491435	1.671139
Marlborough	-0.004645711	0.9308882	3.272825	3.564568	4.096477
New_York	0.782712010	-2.9616338	3.815811	2.849237	4.919106
Oregon	2.744276860	-2.2503736	3.724253	2.957143	4.056568
	aromasTRUE	bodiedTRUE	earthTRUE	forestTRUE	offersTRUE
California	3.6439574	3.1391903	3.634394	4.343958	1.3704178
Casablanca_Valley	5.7297169	2.4463713	3.705557	1.514228	0.1120805
Marlborough	2.7851630	3.7523902	4.307556	4.403705	1.3266324
New_York	2.8690169	3.7222451	4.335826	2.696723	-0.8010808
Oregon	0.9459013	0.8677576	3.575171	2.224639	1.0973870
	raspberryTRUE	smoothTRUE	spiceTRUE	textureTRUE	finishTRUE
California	0.8460102	0.5293772	-0.1996805	0.8533104	1.414024
Casablanca_Valley	2.7819477	1.4035946	1.3082230	-2.8019570	3.185967
Marlborough	-0.7162472	1.2767656	-1.5948409	0.2435642	2.686562
New_York	1.3646539	-1.9712688	0.1666820	-1.7221301	2.646113
Oregon	0.6423684	0.7358623	-1.5175011	-1.7813671	1.682287
	flavorTRUE	fruitTRUE	notesTRUE	sweetTRUE	touchTRUE
California	1.0873866	-1.26884697	2.294055	0.5982191	-1.853642
Casablanca_Valley	3.3747378	-1.22965762	3.837704	1.7436841	-2.219160

Marlborough	-0.4006780	0.02324362	4.152935	-1.0059028	-1.012609	
New_York	0.9511496	-1.81661253	4.251899	1.2714178	-3.005321	
Oregon	1.0969265	0.69160971	2.319266	0.9579421	-1.396591	
	flavorsTRUE	tanninsTRUE	fruityTRUE	strawberryTRUE		
California	-0.3409887	-1.6776045	-2.521548	-0.1487858		
Casablanca_Valley	2.3344046	-2.8408571	-1.884170	-2.6385206		
Marlborough	-0.4535177	-0.8731573	-6.202920	-3.3645528		
New_York	0.4065842	1.4762037	-3.643402	-0.1588614		
Oregon	0.8668873	-1.5327612	-1.850547	-0.5642598		
	cranberryTRUE	darkTRUE	palateTRUE	acidityTRUE	blackTRUE	
California	2.2420927	-0.03308997	2.611196	-1.5886729	0.1379426	
Casablanca_Valley	0.1972379	-1.19743158	3.870187	-0.6427209	-0.2267637	
Marlborough	1.2378438	-0.77904225	2.386708	-5.0288805	0.1630800	
New_York	1.6831065	-3.26534740	3.932216	-1.3184047	1.0773684	
Oregon	1.6529194	-2.07884724	1.542711	-3.3202795	-0.7290467	
	cherryTRUE	colaTRUE	driedTRUE	noseTRUE	softTRUE	juicyTRUE
California	2.045379	4.123931	2.826463	1.846072	-1.585790	-1.7692667
Casablanca_Valley	2.021883	2.891076	2.768651	2.233732	-2.229882	-0.8354553
Marlborough	2.324624	4.676421	2.474280	-0.632064	-2.991824	-8.9713500
New_York	3.874359	2.013664	3.567936	1.368490	-1.724679	-2.5463012
Oregon	1.777277	4.839267	1.723301	-1.104719	-2.775285	-2.3068448
	ripeTRUE	lightTRUE	spicyTRUE	redTRUE	ageTRUE	
California	-0.81774625	-1.0532169	1.743568	-2.00434039	-1.44999380	
Casablanca_Valley	-0.72422165	-2.1729740	2.544577	-0.94327475	-1.21012422	
Marlborough	-1.62295402	-1.7742070	-1.056648	-2.93739351	-1.31354306	
New_York	0.04775652	-2.5605727	-1.492944	-0.07849631	-0.96337152	
Oregon	-1.45044987	-0.9071227	1.322941	-3.25791474	-0.03276851	
	bitTRUE	tightTRUE	cherriesTRUE	coreTRUE	fruitsTRUE	
California	1.967714	-1.467753	1.8029447	-0.9418752	-3.3181590	
Casablanca_Valley	2.878874	-1.278929	-0.6114329	-1.4443096	-1.4400179	
Marlborough	2.707458	-3.119091	3.8473412	-1.9291094	-3.7101744	
New_York	3.113737	-4.263706	2.4750763	1.0239751	-4.5631192	
Oregon	3.145447	-0.885478	2.1828054	-0.3947499	-0.4949466	
	richTRUE	agingTRUE	brightTRUE	characterTRUE		
California	-0.6855453	-4.086215	1.3340623	-2.153439		
Casablanca_Valley	-2.5142322	-2.982069	0.0507180	-2.116107		
Marlborough	-1.4430898	-4.470252	2.0691414	-5.834644		
New_York	-0.3290820	-1.285072	2.9191786	-2.563636		
Oregon	-2.3326986	-1.931577	0.3737944	-2.635048		
	concentratedTRUE	vintageTRUE	complexTRUE	estateTRUE	teaTRUE	
California	1.4620805	-0.1012241	-0.18239771	0.7471917	3.788616	
Casablanca_Valley	0.2357555	0.3845439	-1.67683776	1.1720331	4.272094	
Marlborough	-0.1911979	-0.1456724	0.94191790	-3.2221563	0.257134	

New_York	3.5023071	-2.4667464	0.36091202	0.1426499	1.738348
Oregon	0.2231372	0.3432540	-0.02110645	1.0526559	2.655776
	wildTRUE	firmTRUE	noirTRUE	mediumTRUE	structureTRUE
California	0.36587165	-0.8830966	0.5862657	1.3441821	-2.231292
Casablanca_Valley	-0.08667357	-0.6666827	1.7229977	-1.7538364	-1.002573
Marlborough	-5.39089577	-1.5821823	1.5217391	2.3275299	-3.328536
New_York	-2.01806667	-1.9488939	3.3328402	-0.5876875	-3.888158
Oregon	0.02935259	-2.0077500	0.4596088	0.6494486	-3.676714
	cloveTRUE	timeTRUE	freshTRUE	balancedTRUE	structuredTRUE
California	4.030044	-0.7866540	-0.313762895	0.2123106	-1.897127
Casablanca_Valley	2.424036	-0.1663997	0.548850168	0.5976676	-2.919778
Marlborough	3.132991	-1.3102947	-0.445204013	-2.2028568	-1.448782
New_York	1.889382	-0.9747153	0.004642212	0.1659371	-3.120470
Oregon	2.067396	-1.0542399	-0.183397685	0.2887585	-3.134053
	orangeTRUE	plumTRUE	pomegranateTRUE	cinnamonTRUE	
California	4.145574	-1.00369795	8.030843	2.954537	
Casablanca_Valley	2.615550	2.35394615	7.002037	1.609312	
Marlborough	-1.042343	-1.38485325	7.315746	1.707957	
New_York	2.116361	-0.06534999	3.543369	2.702661	
Oregon	3.372320	-0.66825529	6.067311	2.131381	
	savoryTRUE	pepperTRUE	roseTRUE	points	year
California	4.781396	5.126222	2.9289451	0.84070572	0.011263065
Casablanca_Valley	4.556224	5.480111	0.1823699	-0.63148683	0.004320901
Marlborough	5.557808	4.673293	1.8198124	0.81931570	0.011543528
New_York	6.966426	2.930796	1.8367603	-0.03283032	0.007460148
Oregon	1.685422	5.078627	2.0875089	1.07523317	0.010061348
	year_fold	year_frecent	lprice	price_flow	price_fmed
California	-6.022472	-3.6719643	-3.969975	-1.9204502	-2.3529462
Casablanca_Valley	-4.872445	-0.3158174	-4.158923	-0.3247519	-1.5250940
Marlborough	-3.285649	-0.1375306	-6.252223	-2.5110184	-0.8498381
New_York	-4.203493	-1.9243683	-5.463697	-0.1761111	0.3296439
Oregon	-2.879512	-1.5694764	-4.005269	-2.5947057	-2.0220379
	cost_per_point				
California	-5.062955e-05				
Casablanca_Valley	5.390502e-03				
Marlborough	-3.362881e-04				
New_York	1.765178e-05				
Oregon	1.620567e-04				

Residual Deviance: 3646.722

AIC: 4486.722

```
# Summarize the model
summary(model)
```

Call:

```
nnet::multinom(formula = province ~ ., data = train, trControl = control)
```

Coefficients:

	(Intercept)	bottlingTRUE	earthyTRUE	herbalTRUE	berryTRUE
California	-2.04158110	2.3326692	2.33442850	2.6184997	-1.48538107
Casablanca_Valley	-0.06739038	0.4766752	1.61014621	4.4720869	1.16357544
Marlborough	0.27853449	2.1998386	-0.05894852	2.8450814	-2.26217716
New_York	-0.05014946	-2.1480213	1.15189817	0.2317209	0.02374801
Oregon	1.42782315	1.3700733	0.98981926	2.6651440	-0.54422312
	chocolateTRUE	drinkTRUE	herbTRUE	oakTRUE	tartTRUE
California	-0.166739656	-3.7928243	3.533443	3.913317	3.450648
Casablanca_Valley	3.226693701	-1.0833661	1.330802	4.491435	1.671139
Marlborough	-0.004645711	0.9308882	3.272825	3.564568	4.096477
New_York	0.782712010	-2.9616338	3.815811	2.849237	4.919106
Oregon	2.744276860	-2.2503736	3.724253	2.957143	4.056568
	aromasTRUE	bodiedTRUE	earthTRUE	forestTRUE	offersTRUE
California	3.6439574	3.1391903	3.634394	4.343958	1.3704178
Casablanca_Valley	5.7297169	2.4463713	3.705557	1.514228	0.1120805
Marlborough	2.7851630	3.7523902	4.307556	4.403705	1.3266324
New_York	2.8690169	3.7222451	4.335826	2.696723	-0.8010808
Oregon	0.9459013	0.8677576	3.575171	2.224639	1.0973870
	raspberryTRUE	smoothTRUE	spiceTRUE	textureTRUE	finishTRUE
California	0.8460102	0.5293772	-0.1996805	0.8533104	1.414024
Casablanca_Valley	2.7819477	1.4035946	1.3082230	-2.8019570	3.185967
Marlborough	-0.7162472	1.2767656	-1.5948409	0.2435642	2.686562
New_York	1.3646539	-1.9712688	0.1666820	-1.7221301	2.646113
Oregon	0.6423684	0.7358623	-1.5175011	-1.7813671	1.682287
	flavorTRUE	fruitTRUE	notesTRUE	sweetTRUE	touchTRUE
California	1.0873866	-1.26884697	2.294055	0.5982191	-1.853642
Casablanca_Valley	3.3747378	-1.22965762	3.837704	1.7436841	-2.219160
Marlborough	-0.4006780	0.02324362	4.152935	-1.0059028	-1.012609
New_York	0.9511496	-1.81661253	4.251899	1.2714178	-3.005321
Oregon	1.0969265	0.69160971	2.319266	0.9579421	-1.396591
	flavorsTRUE	tanninsTRUE	fruityTRUE	strawberryTRUE	
California	-0.3409887	-1.6776045	-2.521548	-0.1487858	
Casablanca_Valley	2.3344046	-2.8408571	-1.884170	-2.6385206	
Marlborough	-0.4535177	-0.8731573	-6.202920	-3.3645528	
New_York	0.4065842	1.4762037	-3.643402	-0.1588614	

Oregon	0.8668873	-1.5327612	-1.850547	-0.5642598		
	cranberryTRUE	darkTRUE	palateTRUE	acidityTRUE	blackTRUE	
California	2.2420927	-0.03308997	2.611196	-1.5886729	0.1379426	
Casablanca_Valley	0.1972379	-1.19743158	3.870187	-0.6427209	-0.2267637	
Marlborough	1.2378438	-0.77904225	2.386708	-5.0288805	0.1630800	
New_York	1.6831065	-3.26534740	3.932216	-1.3184047	1.0773684	
Oregon	1.6529194	-2.07884724	1.542711	-3.3202795	-0.7290467	
	cherryTRUE	colaTRUE	driedTRUE	noseTRUE	softTRUE	juicyTRUE
California	2.045379	4.123931	2.826463	1.846072	-1.585790	-1.7692667
Casablanca_Valley	2.021883	2.891076	2.768651	2.233732	-2.229882	-0.8354553
Marlborough	2.324624	4.676421	2.474280	-0.632064	-2.991824	-8.9713500
New_York	3.874359	2.013664	3.567936	1.368490	-1.724679	-2.5463012
Oregon	1.777277	4.839267	1.723301	-1.104719	-2.775285	-2.3068448
	ripeTRUE	lightTRUE	spicyTRUE	redTRUE	ageTRUE	
California	-0.81774625	-1.0532169	1.743568	-2.00434039	-1.44999380	
Casablanca_Valley	-0.72422165	-2.1729740	2.544577	-0.94327475	-1.21012422	
Marlborough	-1.62295402	-1.7742070	-1.056648	-2.93739351	-1.31354306	
New_York	0.04775652	-2.5605727	-1.492944	-0.07849631	-0.96337152	
Oregon	-1.45044987	-0.9071227	1.322941	-3.25791474	-0.03276851	
	bitTRUE	tightTRUE	cherriesTRUE	coreTRUE	fruitsTRUE	
California	1.967714	-1.467753	1.8029447	-0.9418752	-3.3181590	
Casablanca_Valley	2.878874	-1.278929	-0.6114329	-1.4443096	-1.4400179	
Marlborough	2.707458	-3.119091	3.8473412	-1.9291094	-3.7101744	
New_York	3.113737	-4.263706	2.4750763	1.0239751	-4.5631192	
Oregon	3.145447	-0.885478	2.1828054	-0.3947499	-0.4949466	
	richTRUE	agingTRUE	brightTRUE	characterTRUE		
California	-0.6855453	-4.086215	1.3340623	-2.153439		
Casablanca_Valley	-2.5142322	-2.982069	0.0507180	-2.116107		
Marlborough	-1.4430898	-4.470252	2.0691414	-5.834644		
New_York	-0.3290820	-1.285072	2.9191786	-2.563636		
Oregon	-2.3326986	-1.931577	0.3737944	-2.635048		
	concentratedTRUE	vintageTRUE	complexTRUE	estateTRUE	teaTRUE	
California	1.4620805	-0.1012241	-0.18239771	0.7471917	3.788616	
Casablanca_Valley	0.2357555	0.3845439	-1.67683776	1.1720331	4.272094	
Marlborough	-0.1911979	-0.1456724	0.94191790	-3.2221563	0.257134	
New_York	3.5023071	-2.4667464	0.36091202	0.1426499	1.738348	
Oregon	0.2231372	0.3432540	-0.02110645	1.0526559	2.655776	
	wildTRUE	firmTRUE	noirTRUE	mediumTRUE	structureTRUE	
California	0.36587165	-0.8830966	0.5862657	1.3441821	-2.231292	
Casablanca_Valley	-0.08667357	-0.6666827	1.7229977	-1.7538364	-1.002573	
Marlborough	-5.39089577	-1.5821823	1.5217391	2.3275299	-3.328536	
New_York	-2.01806667	-1.9488939	3.3328402	-0.5876875	-3.888158	
Oregon	0.02935259	-2.0077500	0.4596088	0.6494486	-3.676714	

	cloveTRUE	timeTRUE	freshTRUE	balancedTRUE	structuredTRUE
California	4.030044	-0.7866540	-0.313762895	0.2123106	-1.897127
Casablanca_Valley	2.424036	-0.1663997	0.548850168	0.5976676	-2.919778
Marlborough	3.132991	-1.3102947	-0.445204013	-2.2028568	-1.448782
New_York	1.889382	-0.9747153	0.004642212	0.1659371	-3.120470
Oregon	2.067396	-1.0542399	-0.183397685	0.2887585	-3.134053
	orangeTRUE	plumTRUE	pomegranateTRUE	cinnamonTRUE	
California	4.145574	-1.00369795	8.030843	2.954537	
Casablanca_Valley	2.615550	2.35394615	7.002037	1.609312	
Marlborough	-1.042343	-1.38485325	7.315746	1.707957	
New_York	2.116361	-0.06534999	3.543369	2.702661	
Oregon	3.372320	-0.66825529	6.067311	2.131381	
	savoryTRUE	pepperTRUE	roseTRUE	points	year
California	4.781396	5.126222	2.9289451	0.84070572	0.011263065
Casablanca_Valley	4.556224	5.480111	0.1823699	-0.63148683	0.004320901
Marlborough	5.557808	4.673293	1.8198124	0.81931570	0.011543528
New_York	6.966426	2.930796	1.8367603	-0.03283032	0.007460148
Oregon	1.685422	5.078627	2.0875089	1.07523317	0.010061348
	year_fold	year_frecent	lprice	price_flow	price_fmed
California	-6.022472	-3.6719643	-3.969975	-1.9204502	-2.3529462
Casablanca_Valley	-4.872445	-0.3158174	-4.158923	-0.3247519	-1.5250940
Marlborough	-3.285649	-0.1375306	-6.252223	-2.5110184	-0.8498381
New_York	-4.203493	-1.9243683	-5.463697	-0.1761111	0.3296439
Oregon	-2.879512	-1.5694764	-4.005269	-2.5947057	-2.0220379
	cost_per_point				
California	-5.062955e-05				
Casablanca_Valley	5.390502e-03				
Marlborough	-3.362881e-04				
New_York	1.765178e-05				
Oregon	1.620567e-04				

Std. Errors:

	(Intercept)	bottlingTRUE	earthyTRUE	herbalTRUE
California	1.634479e-04	0.052394740	0.087425374	0.0363635639
Casablanca_Valley	1.676385e-05	0.001271790	0.001825742	0.0020801923
Marlborough	5.958074e-05	0.002632244	0.002602181	0.0051086511
New_York	5.066043e-05	0.001044094	0.002409769	0.0005134238
Oregon	2.058716e-04	0.051364363	0.088170261	0.0403089271
	berryTRUE	chocolateTRUE	drinkTRUE	herbTRUE
California	0.064136265	0.0440677466	0.068699301	0.0643802852
Casablanca_Valley	0.002634178	0.0012903933	0.001424085	0.0003903905
Marlborough	0.002072363	0.0032988582	0.018499839	0.0034591942
New_York	0.004873782	0.0006566366	0.004318337	0.0022233629

Oregon	0.069587540	0.0500449153	0.095646979	0.0662974726	
	oakTRUE	tartTRUE	aromasTRUE	bodiedTRUE	earthTRUE
California	0.069268452	0.083538792	0.085418780	0.078350865	0.090175890
Casablanca_Valley	0.002073588	0.000700617	0.004830836	0.002417996	0.001139709
Marlborough	0.004564195	0.007936782	0.010612347	0.018201663	0.005658786
New_York	0.001609032	0.005362455	0.007514962	0.005054519	0.003358451
Oregon	0.069174192	0.085667168	0.072341724	0.058504371	0.089119778
	forestTRUE	offersTRUE	raspberryTRUE	smoothTRUE	
California	0.0071824459	0.086546878	0.071161951	0.0438562050	
Casablanca_Valley	0.0003904089	0.001803994	0.002773308	0.0008240811	
Marlborough	0.0020332319	0.002815781	0.001108713	0.0027082168	
New_York	0.0010685538	0.001030985	0.002891732	0.0007626984	
Oregon	0.0069342564	0.087184574	0.070751962	0.0473971761	
	spiceTRUE	textureTRUE	finishTRUE	flavorTRUE	fruitTRUE
California	0.088299922	0.0583585202	0.060750868	0.082818300	0.053234492
Casablanca_Valley	0.003230738	0.0005398264	0.003058210	0.002061629	0.003174321
Marlborough	0.004517142	0.0092966641	0.009971967	0.003062029	0.016112477
New_York	0.006407289	0.0011794843	0.004784529	0.004054211	0.004280857
Oregon	0.082063163	0.0483610553	0.062781482	0.085813976	0.055922067
	notesTRUE	sweetTRUE	touchTRUE	flavorsTRUE	tanninsTRUE
California	0.088143394	0.020063123	0.072536445	0.054304749	0.0680145240
Casablanca_Valley	0.002303551	0.001237326	0.001574167	0.002499163	0.0009968369
Marlborough	0.008524827	0.001581028	0.003318576	0.010726800	0.0051289070
New_York	0.007581450	0.002198347	0.002021413	0.010414930	0.0094627171
Oregon	0.089069942	0.019489833	0.075883729	0.058922451	0.0741248602
	fruityTRUE	strawberryTRUE	cranberryTRUE	darkTRUE	
California	0.0138605700	0.0739145054	0.083671959	0.075147245	
Casablanca_Valley	0.0007476780	0.0015922780	0.001321819	0.002038516	
Marlborough	0.0006005338	0.0006845995	0.003582770	0.002004659	
New_York	0.0009696406	0.0037714862	0.003717090	0.001362278	
Oregon	0.0151037015	0.0744069349	0.083889781	0.074364459	
	palateTRUE	acidityTRUE	blackTRUE	cherryTRUE	colaTRUE
California	0.070670352	0.075363693	0.066190248	0.055027559	0.0866522572
Casablanca_Valley	0.004570944	0.003628772	0.003442059	0.004737464	0.0013797354
Marlborough	0.007388107	0.003154068	0.015024837	0.011151672	0.0041831094
New_York	0.005789123	0.010127102	0.018242266	0.011356389	0.0009057516
Oregon	0.067990677	0.072719264	0.065577003	0.057321501	0.0860430284
	driedTRUE	noseTRUE	softTRUE	juicyTRUE	ripeTRUE
California	0.051554713	0.072807994	0.0713812467	4.704839e-02	0.078908188
Casablanca_Valley	0.001583625	0.003652440	0.0008256454	1.645655e-03	0.002092487
Marlborough	0.002189518	0.004462424	0.0033216806	2.451663e-06	0.003855315
New_York	0.002549846	0.006267105	0.0029848355	1.459998e-03	0.006538688
Oregon	0.049923015	0.062906101	0.0730203891	4.694523e-02	0.080360227

	lightTRUE	spicyTRUE	redTRUE	ageTRUE	bitTRUE
California	0.070740166	0.084757249	0.070596009	0.0110093766	0.044476202
Casablanca_Valley	0.001826404	0.002553907	0.002887191	0.0002126764	0.003245957
Marlborough	0.004613663	0.001046420	0.004671160	0.0020331523	0.005832009
New_York	0.004575486	0.001448208	0.005285194	0.0015564991	0.004835760
Oregon	0.074254200	0.084435746	0.070642668	0.0150616035	0.048951790
	tightTRUE	cherriesTRUE	coreTRUE	fruitsTRUE	
California	0.0137648471	0.0417276017	0.0168389411	0.0225120544	
Casablanca_Valley	0.0011496407	0.0002554425	0.0008984171	0.0018925371	
Marlborough	0.0007924505	0.0092688314	0.0016806308	0.0002760635	
New_York	0.0004642802	0.0033383315	0.0027109676	0.0005244626	
Oregon	0.0138883309	0.0385912414	0.0163530445	0.0288134622	
	richTRUE	agingTRUE	brightTRUE	characterTRUE	
California	0.032563094	0.0080640195	0.0731548454	0.0123054255	
Casablanca_Valley	0.001127559	0.0002028491	0.0009506553	0.0009984553	
Marlborough	0.005769236	0.0007450727	0.0021157055	0.0004322819	
New_York	0.006174692	0.0034056176	0.0029841075	0.0014413774	
Oregon	0.028669278	0.0114296440	0.0728835819	0.0149504550	
	concentratedTRUE	vintageTRUE	complexTRUE	estateTRUE	
California	0.0254375259	0.0269980872	0.0296893509	0.0475520696	
Casablanca_Valley	0.0006158332	0.0006644211	0.0008293561	0.0004596304	
Marlborough	0.0027885385	0.0025525187	0.0048318241	0.0023128133	
New_York	0.0090976613	0.0001146065	0.0024648898	0.0017080222	
Oregon	0.0249839717	0.0289054228	0.0286374867	0.0489981688	
	teaTRUE	wildTRUE	firmTRUE	noirTRUE	
California	0.040131352	2.629022e-02	0.0194399869	0.077054041	
Casablanca_Valley	0.001164220	7.399204e-04	0.0005288864	0.002159728	
Marlborough	0.001483994	4.972246e-06	0.0030811002	0.011101906	
New_York	0.002166048	1.834868e-03	0.0043680912	0.005146611	
Oregon	0.039307640	2.598851e-02	0.0177498892	0.084000219	
	mediumTRUE	structureTRUE	cloveTRUE	timeTRUE	
California	0.0542876074	0.011799151	0.0145411629	0.019628872	
Casablanca_Valley	0.0006773466	0.001082429	0.0004942463	0.001034494	
Marlborough	0.0173721012	0.001748053	0.0023750159	0.001457243	
New_York	0.0051866886	0.001962556	0.0002785856	0.001207018	
Oregon	0.0414120961	0.013802460	0.0141356492	0.019835518	
	freshTRUE	balancedTRUE	structuredTRUE	orangeTRUE	
California	0.089374464	0.0583193980	0.0075392630	2.863345e-02	
Casablanca_Valley	0.001775204	0.0009750446	0.0003891031	4.827532e-04	
Marlborough	0.002848773	0.0006892599	0.0027826070	1.121624e-05	
New_York	0.004776672	0.0020740649	0.0020204403	2.841666e-04	
Oregon	0.088688224	0.0605854522	0.0075273954	2.854065e-02	
	plumTRUE	pomegranateTRUE	cinnamonTRUE	savoryTRUE	

California	0.085474092	0.0135731389	0.0441136844	0.008312937	
Casablanca_Valley	0.005358577	0.0004627147	0.0008384968	0.001195433	
Marlborough	0.004248676	0.0018608029	0.0029961260	0.004520740	
New_York	0.006522145	0.0004065871	0.0029389604	0.003038904	
Oregon	0.081921878	0.0126719338	0.0412596727	0.006260571	
	pepperTRUE	roseTRUE	points	year	
California	0.0253466800	0.0445321159	0.06241054	1.492347e-04	
Casablanca_Valley	0.0009147281	0.0003362797	0.01742825	1.244470e-04	
Marlborough	0.0016924811	0.0022227267	0.09411294	9.215964e-05	
New_York	0.0010552841	0.0015544688	0.07576594	1.052954e-04	
Oregon	0.0248518904	0.0437173069	0.06119419	1.522296e-04	
	year_fold	year_frecent	lprice	price_flow	price_fmed
California	0.0116773763	0.060203172	0.070489768	0.042693283	0.064024158
Casablanca_Valley	0.0009064389	0.005128182	0.006597021	0.003086840	0.003735262
Marlborough	0.0041831428	0.019696189	0.028132555	0.007248148	0.013349257
New_York	0.0039724815	0.009046872	0.023875117	0.010888515	0.005872906
Oregon	0.0148564855	0.086814898	0.069301519	0.036211798	0.059838905
	cost_per_point				
California	0.0007707637				
Casablanca_Valley	0.0011653763				
Marlborough	0.0011428836				
New_York	0.0012912795				
Oregon	0.0007446660				

Residual Deviance: 3646.722

AIC: 4486.722

```
# Make predictions
preds <- predict(model, type="class", newdata=test)

head(preds)
```

```
[1] Oregon Oregon Oregon Oregon Burgundy Oregon
6 Levels: Burgundy California Casablanca_Valley Marlborough ... Oregon
```

```
postResample(test$province,preds)
```

```
Accuracy      Kappa
0.8804543 0.8145531
```

```
predictors(model)
```

```
[1] "bottling"      "earthy"      "herbal"      "berry"
[5] "chocolate"    "drink"       "herb"        "oak"
[9] "tart"         "aromas"     "bodied"     "earth"
[13] "forest"       "offers"     "raspberry"  "smooth"
[17] "spice"        "texture"    "finish"     "flavor"
[21] "fruit"        "notes"     "sweet"      "touch"
[25] "flavors"      "tannins"    "fruity"     "strawberry"
[29] "cranberry"    "dark"       "palate"     "acidity"
[33] "black"        "cherry"     "cola"       "dried"
[37] "nose"         "soft"       "juicy"      "ripe"
[41] "light"        "spicy"      "red"        "age"
[45] "bit"          "tight"     "cherries"   "core"
[49] "fruits"       "rich"       "aging"      "bright"
[53] "character"    "concentrated" "vintage"    "complex"
[57] "estate"       "tea"        "wild"       "firm"
[61] "noir"         "medium"     "structure"  "clove"
[65] "time"         "fresh"      "balanced"   "structured"
[69] "orange"       "plum"       "pomegranate" "cinnamon"
[73] "savory"       "pepper"     "rose"       "points"
[77] "year"         "year_f"     "lprice"     "price_f"
[81] "cost_per_point"
```

```
varImp(model)%>%
  arrange(desc(Overall))
```

```
Overall
pomegranateTRUE 31.959304191
lprice          23.850087811
savoryTRUE      23.547277352
pepperTRUE      23.289049038
year_fold       21.263571327
earthTRUE       19.558503049
colaTRUE        18.544359354
tartTRUE        18.193937972
oakTRUE         17.775700905
notesTRUE       16.855859307
juicyTRUE       16.429217979
fruityTRUE      16.102586395
```

aromasTRUE	15.973755418
herbTRUE	15.677134439
characterTRUE	15.302873079
forestTRUE	15.183254031
agingTRUE	14.755185444
palateTRUE	14.343019231
structureTRUE	14.127273001
bodiedTRUE	13.927954473
bitTRUE	13.813231368
cloveTRUE	13.543849335
fruitsTRUE	13.526417204
driedTRUE	13.360631088
orangeTRUE	13.292147137
herbalTRUE	12.832532875
teaTRUE	12.711967210
structuredTRUE	12.520209495
cherryTRUE	12.043522523
acidityTRUE	11.898958413
finishTRUE	11.614953459
softTRUE	11.307460565
cinnamonTRUE	11.105847688
drinkTRUE	11.019086079
tightTRUE	11.014957054
cherriesTRUE	10.919600554
touchTRUE	9.487323060
redTRUE	9.221419710
roseTRUE	8.855396655
bottlingTRUE	8.527277577
lightTRUE	8.468093337
tanninsTRUE	8.400583928
spicyTRUE	8.160678692
wildTRUE	7.890860253
noirTRUE	7.623451525
year_frecent	7.619157019
price_flow	7.527037338
textureTRUE	7.402328892
darkTRUE	7.353758444
richTRUE	7.304647876
noseTRUE	7.185077581
firmTRUE	7.088605393
price_fmed	7.079560067
cranberryTRUE	7.013200314
chocolateTRUE	6.925067938

```

flavorTRUE      6.910878338
strawberryTRUE  6.874980417
brightTRUE      6.746894715
mediumTRUE      6.662684592
raspberryTRUE   6.351227509
estateTRUE      6.336686848
earthyTRUE      6.145240673
smoothTRUE      5.916868534
coreTRUE        5.734019219
concentratedTRUE 5.614478136
sweetTRUE       5.577165824
berryTRUE       5.479104796
plumTRUE        5.476102631
fruitTRUE       5.029970462
ageTRUE         4.969801105
spiceTRUE       4.786927595
offersTRUE      4.707598594
ripeTRUE        4.663128312
flavorsTRUE     4.402382535
timeTRUE        4.292303707
balancedTRUE    3.467530639
vintageTRUE     3.441440785
points          3.399571733
complexTRUE     3.183171843
blackTRUE       2.334201500
freshTRUE       1.495856974
year            0.044648990
cost_per_point  0.005957128

```

```
confusionMatrix(predict(model, test),factor(test$province))
```

Confusion Matrix and Statistics

	Reference				
Prediction	Burgundy	California	Casablanca_Valley	Marlborough	New_York
Burgundy	221	7	0	1	0
California	2	720	4	2	7
Casablanca_Valley	1	1	20	0	0
Marlborough	0	2	0	32	0
New_York	0	7	0	0	17
Oregon	14	54	2	10	2

	Reference
Prediction	Oregon
Burgundy	10
California	70
Casablanca_Valley	1
Marlborough	3
New_York	0
Oregon	463

Overall Statistics

Accuracy : 0.8805
 95% CI : (0.8639, 0.8956)
 No Information Rate : 0.4728
 P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.8146

Mcnemar's Test P-Value : NA

Statistics by Class:

	Class: Burgundy	Class: California	Class: Casablanca_Valley
Sensitivity	0.9286	0.9102	0.76923
Specificity	0.9875	0.9036	0.99818
Pos Pred Value	0.9247	0.8944	0.86957
Neg Pred Value	0.9881	0.9182	0.99636
Prevalence	0.1423	0.4728	0.01554
Detection Rate	0.1321	0.4304	0.01195
Detection Prevalence	0.1429	0.4812	0.01375
Balanced Accuracy	0.9580	0.9069	0.88370
	Class: Marlborough	Class: New_York	Class: Oregon
Sensitivity	0.71111	0.65385	0.8464
Specificity	0.99693	0.99575	0.9272
Pos Pred Value	0.86486	0.70833	0.8495
Neg Pred Value	0.99205	0.99454	0.9255
Prevalence	0.02690	0.01554	0.3270
Detection Rate	0.01913	0.01016	0.2767
Detection Prevalence	0.02212	0.01435	0.3258
Balanced Accuracy	0.85402	0.82480	0.8868