Seasonal investigation of beers (primarily focused on Summer)

|  |  |  |  |
| --- | --- | --- | --- |
| Season | Months | Proportion of ratings | Mean overall rating |
| Winter | Dec-Feb | 26.32% | 3.8306 |
| Spring | Mar-May | 23.99% | 3.8311 |
| Summer | June-Aug | 24.12% | 3.8120 |
| Autumn | Sept-Nov | 25.56% | 3.8152 |

10 most commonly drunk beer styles in summer (Summer = June, July, Aug)

> top.summer.names

[1] "American IPA" "American Double / Imperial IPA" "American Pale Ale (APA)"

[4] "Russian Imperial Stout" "American Double / Imperial Stout" "American Amber / Red Ale"

[7] "American Porter" "Saison / Farmhouse Ale" "Fruit / Vegetable Beer"

[10] "Hefeweizen"

10 beers which have the biggest % of their ratings in summer, compared to the rest of the year. Compare to a 0.25 baseline if each season received an equal proportion of ratings

> sort(tapply(summer.data$rating,summer.data$beerStyle,length)/tapply(data.scaled$rating,

+ data.scaled$beerStyle, length), decreasing=T)[1:10]

Berliner Weissbier Kvass American Pale Wheat Ale KÃ¶lsch

0.4156588 0.3788546 0.3725750 0.3690430

American Blonde Ale Roggenbier Hefeweizen Cream Ale

0.3540910 0.3461538 0.3361570 0.3207789

Sahti Chile Beer

0.3159478 0.3133971

Conversely these have the smallest % of their ratings in summer. As we might expect this includes Winter Warmer etc. Compare to a 0.25 baseline for ‘non-seasonal’ beers

> sort(tapply(summer.data$rating,summer.data$beerStyle,length)/tapply(data.scaled$rating,

+ data.scaled$beerStyle, length), decreasing=F)[1:10]

Winter Warmer Pumpkin Ale MÃ¤rzen / Oktoberfest

0.05645080 0.07725176 0.11636584

Bock BiÃ¨re de Champagne / BiÃ¨re Brut English Strong Ale

0.14564273 0.15850144 0.18972081

Happoshu Belgian Strong Dark Ale American Stout

0.18987342 0.19309586 0.19591400

Oatmeal Stout

0.19679877

Only using the summer data our principal components do not really change compared to the year as a whole. Practically identical.

Rotation:

PC1 PC2 PC3 PC4 PC5 PC6

stand.ABV -0.2105055 -0.92817898 0.20627005 -0.14501813 0.1501811 0.08966175

stand.Appear -0.3898328 -0.06462486 -0.89200464 -0.11227839 0.1867392 -0.02652919

stand.Aroma -0.4387005 -0.05390432 0.02093578 0.81298656 -0.3160801 0.20819247

stand.Pal -0.4452510 0.09613989 0.10371623 -0.52523807 -0.7094808 0.05013558

stand.Taste -0.4695004 0.12334268 0.26728965 0.05031616 0.2573563 -0.79003050

stand.rating -0.4388401 0.32702512 0.28132057 -0.16434537 0.5225631 0.56679018

**PCA with SEASON**

> season.pca

Standard deviations:

[1] 1.9161378 1.0011772 0.9568514 0.7409621 0.6261280 0.5326309 0.4309740

Rotation:

PC1 PC2 PC3 PC4 PC5 PC6 PC7

stand.ABV 0.21419872 0.152324889 -0.91874903 0.188940500 -0.139408501 0.155054713 -0.0874130779

stand.Appear 0.38477220 0.009520219 -0.04477736 -0.893459971 -0.124565853 0.187505622 0.0301492199

stand.Aroma 0.43672607 0.010058505 -0.04799971 0.007765410 0.821927489 -0.290274289 -0.2167508503

stand.Pal 0.44838109 -0.022059836 0.08461393 0.105794623 -0.506695515 -0.720091514 -0.0695960562

stand.Taste 0.47063569 -0.022859249 0.11735571 0.266246081 0.048038831 0.235814641 0.7971222536

stand.rating 0.43904626 -0.058806204 0.32078203 0.289593267 -0.174345870 0.531410490 -0.5515625517

scale.season -0.00586682 -0.985970572 -0.16660806 -0.005169894 0.006266101 0.001752976 0.0005483036

> summary(season.pca)

Importance of components:

PC1 PC2 PC3 PC4 PC5 PC6 PC7

Standard deviation 1.9161 1.0012 0.9569 0.74096 0.62613 0.53263 0.43097

Proportion of Variance 0.5245 0.1432 0.1308 0.07843 0.05601 0.04053 0.02653

Cumulative Proportion 0.5245 0.6677 0.7985 0.87693 0.93294 0.97347 1.00000

> eigen(cov(data.scaled[,c(19:24,32)]))

$values

[1] 3.6715841 1.0023558 0.9155646 0.5490249 0.3920363 0.2836957 0.1857386

$vectors

[,1] [,2] [,3] [,4] [,5] [,6] [,7]

[1,] -0.21419872 -0.152324889 0.91874903 -0.188940500 -0.139408501 0.155054713 0.0874130779

[2,] -0.38477220 -0.009520219 0.04477736 0.893459971 -0.124565853 0.187505622 -0.0301492199

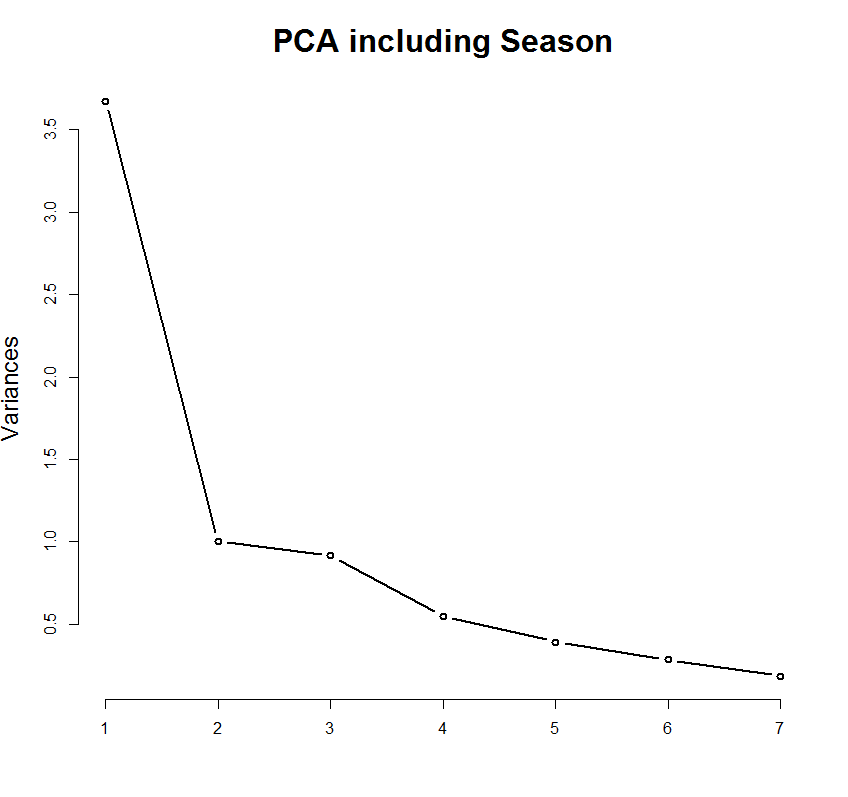
[3,] -0.43672607 -0.010058505 0.04799971 -0.007765410 0.821927489 -0.290274289 0.2167508503

[4,] -0.44838109 0.022059836 -0.08461393 -0.105794623 -0.506695515 -0.720091514 0.0695960562

[5,] -0.47063569 0.022859249 -0.11735571 -0.266246081 0.048038831 0.235814641 -0.7971222536

[6,] -0.43904626 0.058806204 -0.32078203 -0.289593267 -0.174345870 0.531410490 0.5515625517

[7,] 0.00586682 0.985970572 0.16660806 0.005169894 0.006266101 0.001752976 -0.0005483036



**PCA with MONTH**

> month.pca

Standard deviations:

[1] 1.9161670 1.0007231 0.9573034 0.7409659 0.6260889 0.5326113 0.4309700

Rotation:

PC1 PC2 PC3 PC4 PC5 PC6 PC7

stand.ABV 0.214206374 -0.126927168 0.92246481 0.189229544 -0.13972737 -0.155174947 -0.087481231

stand.Appear 0.384760683 -0.001840357 0.04613052 -0.893449624 -0.12450879 -0.187525562 0.030121410

stand.Aroma 0.436695813 0.007195443 0.05077529 0.007855757 0.82170344 0.290614104 -0.216684229

stand.Pal 0.448378979 0.013659454 -0.08625278 0.105832654 -0.50703392 0.719866278 -0.069561239

stand.Taste 0.470624353 0.020052856 -0.11794496 0.266226980 0.04807948 -0.235898833 0.797096535

stand.rating 0.439049398 0.040641855 -0.32379508 0.289458818 -0.17396102 -0.531422860 -0.551617183

scale.month -0.008732005 0.990754195 0.13475424 0.003804382 -0.01094629 -0.005689253 -0.002123957

> summary(month.pca)

Importance of components:

PC1 PC2 PC3 PC4 PC5 PC6 PC7

Standard deviation 1.9162 1.0007 0.9573 0.74097 0.6261 0.53261 0.43097

Proportion of Variance 0.5245 0.1431 0.1309 0.07843 0.0560 0.04052 0.02653

Cumulative Proportion 0.5245 0.6676 0.7985 0.87694 0.9329 0.97347 1.00000

> eigen(cov(data.scaled[,c(19:24,33)]))

$values

[1] 3.6716959 1.0014467 0.9164298 0.5490304 0.3919873 0.2836747 0.1857352

$vectors

[,1] [,2] [,3] [,4] [,5] [,6] [,7]

[1,] -0.214206374 -0.126927168 0.92246481 -0.189229544 -0.13972737 0.155174947 0.087481231

[2,] -0.384760683 -0.001840357 0.04613052 0.893449624 -0.12450879 0.187525562 -0.030121410

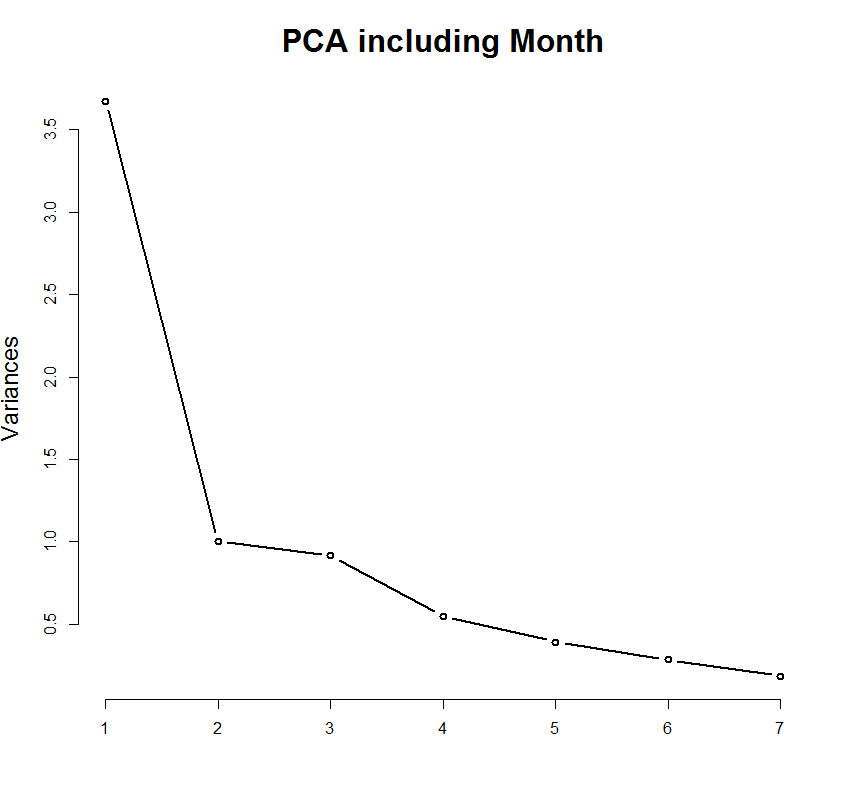
[3,] -0.436695813 0.007195443 0.05077529 -0.007855757 0.82170344 -0.290614104 0.216684229

[4,] -0.448378979 0.013659454 -0.08625278 -0.105832654 -0.50703392 -0.719866278 0.069561239

[5,] -0.470624353 0.020052856 -0.11794496 -0.266226980 0.04807948 0.235898833 -0.797096535

[6,] -0.439049398 0.040641855 -0.32379508 -0.289458818 -0.17396102 0.531422860 0.551617183

[7,] 0.008732005 0.990754195 0.13475424 -0.003804382 -0.01094629 0.005689253 0.002123957



> tapply(data.scaled$rating, data.scaled$month, function(x){

+ (-0.083333+length(x)/length(data.scaled$rating))\*100

+ })

1 2 3 4 5 6 7 8 9

0.67153857 -0.62794314 -0.08525713 -0.63724936 -0.28341330 -0.76953074 -0.28859820 0.17983740 -0.19673244

10 11 12

0.22690032 0.53048276 1.28036526

Percentage of beer ratings per calendar month – expressed in terms of change from expected (100%/12 = 8.33333) ie. In Decemeber there are 8.33333+1.2804 percent of the total yearly ratings

IN BOTH CASES PC 2 BECOMES DOMINATED BY TIME INFORMATION (MONTH/SEASON). Also the eigenvalue is > 1 in both cases.

Is there an issue using month and season in a PCA as they are cyclical and pca ignores this. Adjusted so seasons run 1= spring 2= summer 3= autumn 4= winter

**PCA with time may be of more use**

> time.pca

Standard deviations:

[1] 1.9178038 1.0408539 0.9101234 0.7409555 0.6261143 0.5326329 0.4308075

Rotation:

PC1 PC2 PC3 PC4 PC5 PC6 PC7

stand.ABV 0.2159706 0.545742150 0.75288099 -0.191417872 0.141434858 -0.1549398705 0.08967406

stand.Appear 0.3843691 0.007842268 0.05269474 0.893205055 0.124747125 -0.1874845107 -0.02991007

stand.Aroma 0.4362791 0.008831073 0.05468102 -0.008083303 -0.821724277 0.2903164909 0.21682422

stand.Pal 0.4477194 -0.065813672 -0.06290151 -0.105659422 0.506677931 0.7200684308 0.06981615

stand.Taste 0.4697052 -0.096871839 -0.07682458 -0.266253273 -0.047775336 -0.2357845024 -0.79707169

stand.rating 0.4379309 -0.203943031 -0.25882139 -0.288738747 0.173571941 -0.5314757123 0.55106399

scale.time 0.0491786 0.804187580 -0.59207799 0.007335226 -0.008445192 -0.0004695207 -0.01349489

> summary(time.pca)

Importance of components:

PC1 PC2 PC3 PC4 PC5 PC6 PC7

Standard deviation 1.9178 1.0409 0.9101 0.74096 0.6261 0.53263 0.43081

Proportion of Variance 0.5254 0.1548 0.1183 0.07843 0.0560 0.04053 0.02651

Cumulative Proportion 0.5254 0.6802 0.7985 0.87696 0.9330 0.97349 1.00000

> eigen(cov(data.scaled[,c(19:24,34)]))

$values

[1] 3.6779715 1.0833768 0.8283246 0.5490151 0.3920191 0.2836978 0.1855951

$vectors

[,1] [,2] [,3] [,4] [,5] [,6] [,7]

[1,] -0.2159706 0.545742150 0.75288099 -0.191417872 0.141434858 -0.1549398705 -0.08967406

[2,] -0.3843691 0.007842268 0.05269474 0.893205055 0.124747125 -0.1874845107 0.02991007

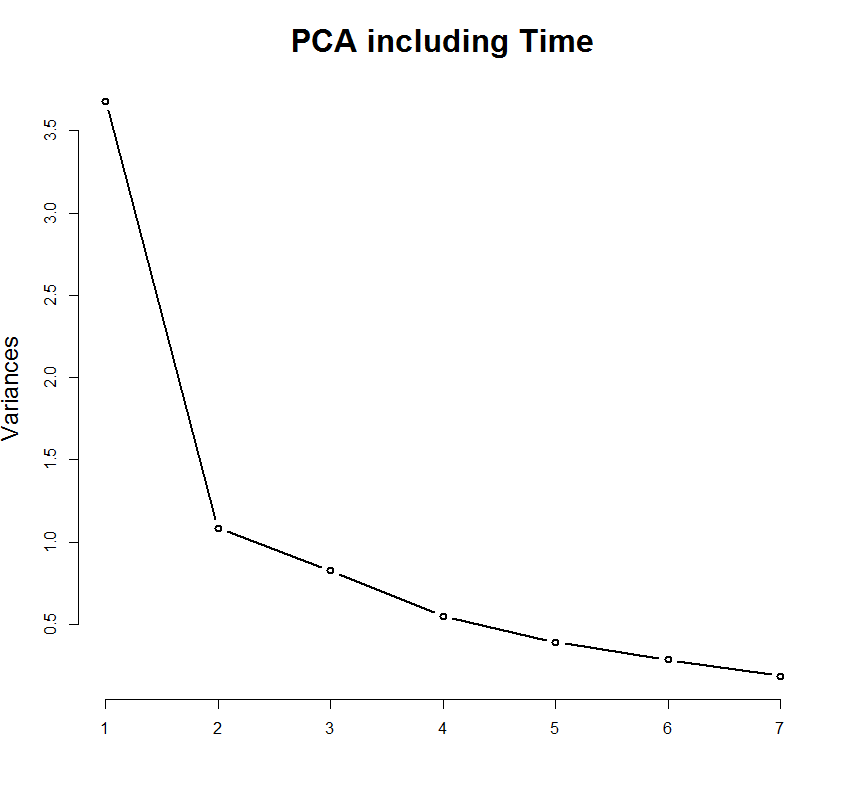
[3,] -0.4362791 0.008831073 0.05468102 -0.008083303 -0.821724277 0.2903164909 -0.21682422

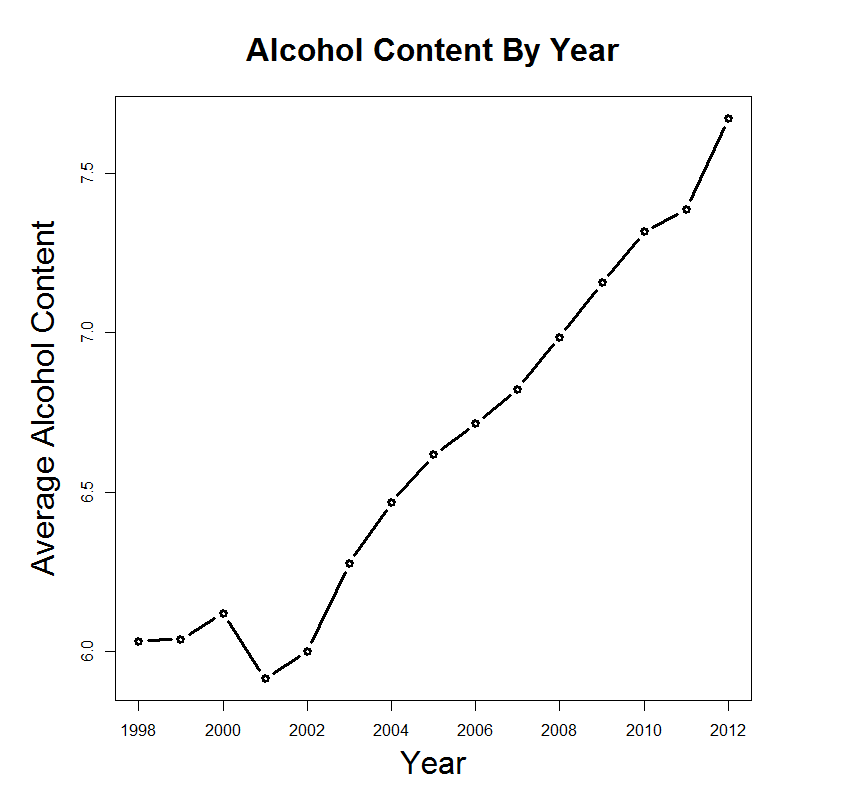
[4,] -0.4477194 -0.065813672 -0.06290151 -0.105659422 0.506677931 0.7200684308 -0.06981615

[5,] -0.4697052 -0.096871839 -0.07682458 -0.266253273 -0.047775336 -0.2357845024 0.79707169

[6,] -0.4379309 -0.203943031 -0.25882139 -0.288738747 0.173571941 -0.5314757123 -0.55106399

[7,] -0.0491786 0.804187580 -0.59207799 0.007335226 -0.008445192 -0.0004695207 0.01349489





> tapply(data.predict$ABV, data.predict$year, mean)

1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

6.032083 6.038788 6.120745 5.915547 6.000489 6.276917 6.468032 6.616829 6.716165 6.823196 6.985672 7.157001 7.315950

2011 2012

7.385910 7.671976

**Could this explain PC2? Or PC3?**