Class 09 Halloween Candy Mini Proj

Dylan Mullaney (A16869792)

Today we will examine data from 538 on common Halloween candy. In this particular project we will use ggplot, dplyr, and PCA to make sense of the multivariable dataset.

Importing candy data

For a txt file, could use read.table

```
url <- "https://raw.githubusercontent.com/fivethirtyeight/data/master/candy-power-ranking/ca
candy <- read.csv (url, row.names=1)
head(candy)
```

```
             chocolate fruity caramel peanutyalmondy nougat crispedricewafer
100 Grand            1      0       1              0      0                1
3 Musketeers         1      0       0              0      1                0
One dime             0      0       0              0      0                0
One quarter          0      0       0              0      0                0
Air Heads            0      1       0              0      0                0
Almond Joy           1      0       0              1      0                0
             hard bar pluribus sugarpercent pricepercent winpercent
100 Grand       0   1        0        0.732        0.860   66.97173
3 Musketeers    0   1        0        0.604        0.511   67.60294
One dime        0   0        0        0.011        0.116   32.26109
One quarter     0   0        0        0.011        0.511   46.11650
Air Heads       0   0        0        0.906        0.511   52.34146
Almond Joy      0   1        0        0.465        0.767   50.34755
```

> Q1. How many different candy types are in this dataset?

```
nrow(candy)
```

```
[1] 85
```

> Q2. Q2. How many fruity candy types are in the dataset?

```
sum(candy$fruity)
```

```
[1] 38
```

```
#This is a count value, not a percent
```

Winpercent refers to the likelihood a candy will be chosen as the winner.

> Q3. What is your favorite candy in the dataset and what is it's winpercent value?

```
candy["Haribo Happy Cola",] $winpercent
```

```
[1] 34.15896
```

```
#This is a percent value, not a count
```

> Q4. What is the winpercent value for "Kit Kat"?

```
candy["Kit Kat",] $winpercent
```

```
[1] 76.7686
```

> Q5. What is the winpercent value for "Tootsie Roll Snack Bars"?

```
candy["Tootsie Roll Snack Bars",] $winpercent
```

```
[1] 49.6535
```

How many candies are chocolate?

```
sum(candy$chocolate)
```

```
[1] 37
```

> Q6. Is there any variable/column that looks to be on a different scale to the majority of the other columns in the dataset?

Winpercent appears to be on different scale.

```r
#Install package in R brain
library("skimr")
skim(candy)
```

Table 1: Data summary

Name	candy
Number of rows	85
Number of columns	12
________________________	
Column type frequency:	
numeric	12
________________________	
Group variables	None

Variable type: numeric

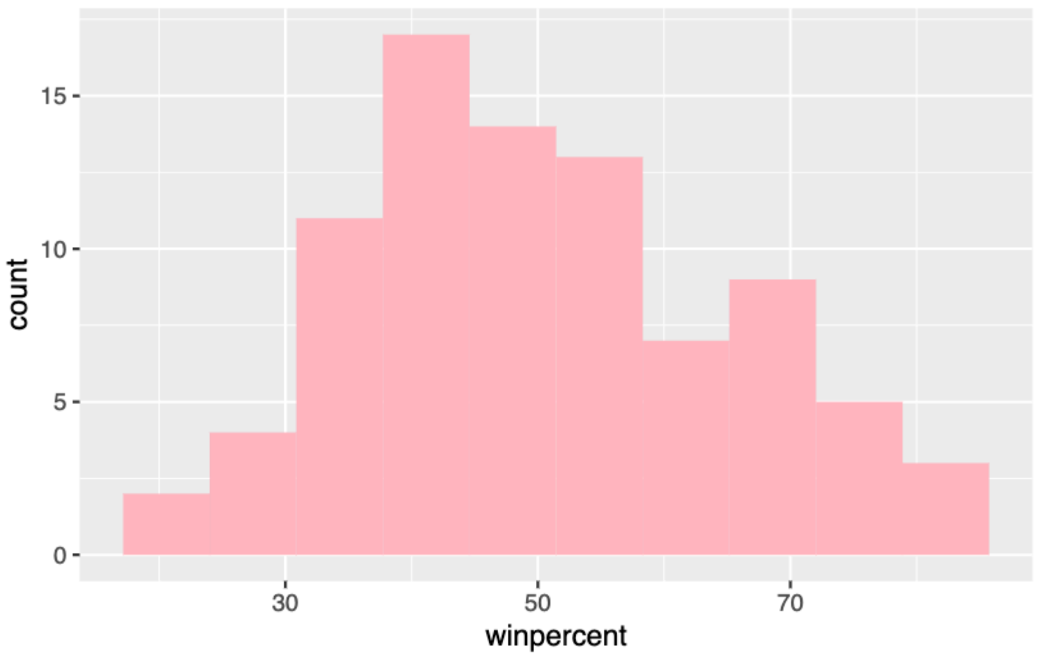
skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
chocolate	0	1	0.44	0.50	0.00	0.00	0.00	1.00	1.00	
fruity	0	1	0.45	0.50	0.00	0.00	0.00	1.00	1.00	
caramel	0	1	0.16	0.37	0.00	0.00	0.00	0.00	1.00	
peanutyalmondy	0	1	0.16	0.37	0.00	0.00	0.00	0.00	1.00	
nougat	0	1	0.08	0.28	0.00	0.00	0.00	0.00	1.00	
crispedricewafer	0	1	0.08	0.28	0.00	0.00	0.00	0.00	1.00	
hard	0	1	0.18	0.38	0.00	0.00	0.00	0.00	1.00	
bar	0	1	0.25	0.43	0.00	0.00	0.00	0.00	1.00	
pluribus	0	1	0.52	0.50	0.00	0.00	1.00	1.00	1.00	
sugarpercent	0	1	0.48	0.28	0.01	0.22	0.47	0.73	0.99	
pricepercent	0	1	0.47	0.29	0.01	0.26	0.47	0.65	0.98	
winpercent	0	1	50.32	14.71	22.45	39.14	47.83	59.86	84.18	

Q7. What do you think a zero and one represent for the candy$chocolate column?

This represents a candy that doesn’t match this identity, ie it isn’t chocolate.

Q8. Plot a histogram of winpercent values

```r
library(ggplot2)

ggplot (candy) +
       aes(winpercent) +
       geom_histogram(bins=10, fill="lightpink")
```



Q9. Is the distribution of winpercent values symmetrical?

No, based on the shape of the histogram.

Q10. Is the center of the distribution above or below 50%?

It appears to be below 50%. Mean is above 50, but the median is below.

```
summary(candy$winpercent)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  22.45   39.14   47.83   50.32   59.86   84.18 
```

Q11. On average is chocolate candy higher or lower ranked than fruit candy?

-Step 1: find all chocolate candy -Step 2: find their winpercent values -Step 3: summarize these values (mean, median) -Step 4: repeat with fruit candies -Step 5: compare the two summart values

Step 1

```
choc.inds <- candy$chocolate == 1
candy[choc.inds,]
```

```
                            chocolate fruity caramel peanutyalmondy nougat
100 Grand                           1      0       1              0      0
3 Musketeers                        1      0       0              0      1
Almond Joy                          1      0       0              1      0
Baby Ruth                           1      0       1              1      1
Charleston Chew                     1      0       0              0      1
Hershey's Kisses                    1      0       0              0      0
Hershey's Krackel                   1      0       0              0      0
Hershey's Milk Chocolate            1      0       0              0      0
Hershey's Special Dark              1      0       0              0      0
Junior Mints                        1      0       0              0      0
Kit Kat                             1      0       0              0      0
Peanut butter M&M's                 1      0       0              1      0
M&M's                               1      0       0              0      0
Milk Duds                           1      0       1              0      0
Milky Way                           1      0       1              0      1
Milky Way Midnight                  1      0       1              0      1
Milky Way Simply Caramel            1      0       1              0      0
Mounds                              1      0       0              0      0
Mr Good Bar                         1      0       0              1      0
Nestle Butterfinger                 1      0       0              1      0
Nestle Crunch                       1      0       0              0      0
Peanut M&Ms                         1      0       0              1      0
Reese's Miniatures                  1      0       0              1      0
Reese's Peanut Butter cup           1      0       0              1      0
Reese's pieces                      1      0       0              1      0
Reese's stuffed with pieces         1      0       0              1      0
Rolo                                1      0       1              0      0
Sixlets                             1      0       0              0      0
Nestle Smarties                     1      0       0              0      0
Snickers                            1      0       1              1      1
Snickers Crisper                    1      0       1              1      0
Tootsie Pop                         1      1       0              0      0
Tootsie Roll Juniors                1      0       0              0      0
Tootsie Roll Midgies                1      0       0              0      0
Tootsie Roll Snack Bars             1      0       0              0      0
Twix                                1      0       1              0      0
Whoppers                            1      0       0              0      0
                            crispedricewafer hard bar pluribus sugarpercent
100 Grand                                  1    0   1        0        0.732
3 Musketeers                               0    0   1        0        0.604
Almond Joy                                 0    0   1        0        0.465
Baby Ruth                                  0    0   1        0        0.604
```

```
Charleston Chew                            0    0   1        0        0.604
Hershey's Kisses                           0    0   0        1        0.127
Hershey's Krackel                          1    0   1        0        0.430
Hershey's Milk Chocolate                   0    0   1        0        0.430
Hershey's Special Dark                     0    0   1        0        0.430
Junior Mints                               0    0   0        1        0.197
Kit Kat                                    1    0   1        0        0.313
Peanut butter M&M's                        0    0   0        1        0.825
M&M's                                      0    0   0        1        0.825
Milk Duds                                  0    0   0        1        0.302
Milky Way                                  0    0   1        0        0.604
Milky Way Midnight                         0    0   1        0        0.732
Milky Way Simply Caramel                   0    0   1        0        0.965
Mounds                                     0    0   1        0        0.313
Mr Good Bar                                0    0   1        0        0.313
Nestle Butterfinger                        0    0   1        0        0.604
Nestle Crunch                              1    0   1        0        0.313
Peanut M&Ms                                0    0   0        1        0.593
Reese's Miniatures                         0    0   0        0        0.034
Reese's Peanut Butter cup                  0    0   0        0        0.720
Reese's pieces                             0    0   0        1        0.406
Reese's stuffed with pieces                0    0   0        0        0.988
Rolo                                       0    0   0        1        0.860
Sixlets                                    0    0   0        1        0.220
Nestle Smarties                            0    0   0        1        0.267
Snickers                                   0    0   1        0        0.546
Snickers Crisper                           1    0   1        0        0.604
Tootsie Pop                                0    1   0        0        0.604
Tootsie Roll Juniors                       0    0   0        0        0.313
Tootsie Roll Midgies                       0    0   0        1        0.174
Tootsie Roll Snack Bars                    0    0   1        0        0.465
Twix                                       1    0   1        0        0.546
Whoppers                                   1    0   0        1        0.872
                            pricepercent winpercent
100 Grand                          0.860   66.97173
3 Musketeers                       0.511   67.60294
Almond Joy                         0.767   50.34755
Baby Ruth                          0.767   56.91455
Charleston Chew                    0.511   38.97504
Hershey's Kisses                   0.093   55.37545
Hershey's Krackel                  0.918   62.28448
Hershey's Milk Chocolate           0.918   56.49050
Hershey's Special Dark             0.918   59.23612
```

```
Junior Mints                       0.511   57.21925
Kit Kat                            0.511   76.76860
Peanut butter M&M's                0.651   71.46505
M&M's                              0.651   66.57458
Milk Duds                          0.511   55.06407
Milky Way                          0.651   73.09956
Milky Way Midnight                 0.441   60.80070
Milky Way Simply Caramel           0.860   64.35334
Mounds                             0.860   47.82975
Mr Good Bar                        0.918   54.52645
Nestle Butterfinger                0.767   70.73564
Nestle Crunch                      0.767   66.47068
Peanut M&Ms                        0.651   69.48379
Reese's Miniatures                 0.279   81.86626
Reese's Peanut Butter cup          0.651   84.18029
Reese's pieces                     0.651   73.43499
Reese's stuffed with pieces        0.651   72.88790
Rolo                               0.860   65.71629
Sixlets                            0.081   34.72200
Nestle Smarties                    0.976   37.88719
Snickers                           0.651   76.67378
Snickers Crisper                   0.651   59.52925
Tootsie Pop                        0.325   48.98265
Tootsie Roll Juniors               0.511   43.06890
Tootsie Roll Midgies               0.011   45.73675
Tootsie Roll Snack Bars            0.325   49.65350
Twix                               0.906   81.64291
Whoppers                           0.848   49.52411
```

Step 2

```r
choc.win <- candy[choc.inds,]$winpercent
choc.win
```

```
 [1] 66.97173 67.60294 50.34755 56.91455 38.97504 55.37545 62.28448 56.49050
 [9] 59.23612 57.21925 76.76860 71.46505 66.57458 55.06407 73.09956 60.80070
[17] 64.35334 47.82975 54.52645 70.73564 66.47068 69.48379 81.86626 84.18029
[25] 73.43499 72.88790 65.71629 34.72200 37.88719 76.67378 59.52925 48.98265
[33] 43.06890 45.73675 49.65350 81.64291 49.52411
```

Step 3

```r
choc.mean <- mean(choc.win)
choc.mean
```

```
[1] 60.92153
```

Step 4

```r
fruit.inds <- candy$fruity == 1
candy[fruit.inds,]
```

```
                            chocolate fruity caramel peanutyalmondy nougat
Air Heads                           0      1       0             0      0
Caramel Apple Pops                  0      1       1             0      0
Chewey Lemonhead Fruit Mix          0      1       0             0      0
Chiclets                            0      1       0             0      0
Dots                                0      1       0             0      0
Dum Dums                            0      1       0             0      0
Fruit Chews                         0      1       0             0      0
Fun Dip                             0      1       0             0      0
Gobstopper                          0      1       0             0      0
Haribo Gold Bears                   0      1       0             0      0
Haribo Sour Bears                   0      1       0             0      0
Haribo Twin Snakes                  0      1       0             0      0
Jawbusters                          0      1       0             0      0
Laffy Taffy                         0      1       0             0      0
Lemonhead                           0      1       0             0      0
Lifesavers big ring gummies         0      1       0             0      0
Mike & Ike                          0      1       0             0      0
Nerds                               0      1       0             0      0
Nik L Nip                           0      1       0             0      0
Now & Later                         0      1       0             0      0
Pop Rocks                           0      1       0             0      0
Red vines                           0      1       0             0      0
Ring pop                            0      1       0             0      0
Runts                               0      1       0             0      0
Skittles original                   0      1       0             0      0
Skittles wildberry                  0      1       0             0      0
Smarties candy                      0      1       0             0      0
Sour Patch Kids                     0      1       0             0      0
Sour Patch Tricksters               0      1       0             0      0
Starburst                           0      1       0             0      0
```

```
Strawberry bon bons                 0      1       0              0      0
Super Bubble                        0      1       0              0      0
Swedish Fish                        0      1       0              0      0
Tootsie Pop                         1      1       0              0      0
Trolli Sour Bites                   0      1       0              0      0
Twizzlers                           0      1       0              0      0
Warheads                            0      1       0              0      0
Welch's Fruit Snacks                0      1       0              0      0
                            crispedricewafer hard bar pluribus sugarpercent
Air Heads                                  0    0   0        0        0.906
Caramel Apple Pops                         0    0   0        0        0.604
Chewey Lemonhead Fruit Mix                 0    0   0        1        0.732
Chiclets                                   0    0   0        1        0.046
Dots                                       0    0   0        1        0.732
Dum Dums                                   0    1   0        0        0.732
Fruit Chews                                0    0   0        1        0.127
Fun Dip                                    0    1   0        0        0.732
Gobstopper                                 0    1   0        1        0.906
Haribo Gold Bears                          0    0   0        1        0.465
Haribo Sour Bears                          0    0   0        1        0.465
Haribo Twin Snakes                         0    0   0        1        0.465
Jawbusters                                 0    1   0        1        0.093
Laffy Taffy                                0    0   0        0        0.220
Lemonhead                                  0    1   0        0        0.046
Lifesavers big ring gummies                0    0   0        0        0.267
Mike & Ike                                 0    0   0        1        0.872
Nerds                                      0    1   0        1        0.848
Nik L Nip                                  0    0   0        1        0.197
Now & Later                                0    0   0        1        0.220
Pop Rocks                                  0    1   0        1        0.604
Red vines                                  0    0   0        1        0.581
Ring pop                                   0    1   0        0        0.732
Runts                                      0    1   0        1        0.872
Skittles original                          0    0   0        1        0.941
Skittles wildberry                         0    0   0        1        0.941
Smarties candy                             0    1   0        1        0.267
Sour Patch Kids                            0    0   0        1        0.069
Sour Patch Tricksters                      0    0   0        1        0.069
Starburst                                  0    0   0        1        0.151
Strawberry bon bons                        0    1   0        1        0.569
Super Bubble                               0    0   0        0        0.162
Swedish Fish                               0    0   0        1        0.604
Tootsie Pop                                0    1   0        0        0.604
```

```
Trolli Sour Bites                          0    0   0        1        0.313
Twizzlers                                  0    0   0        0        0.220
Warheads                                   0    1   0        0        0.093
Welch's Fruit Snacks                       0    0   0        1        0.313
                            pricepercent winpercent
Air Heads                          0.511   52.34146
Caramel Apple Pops                 0.325   34.51768
Chewey Lemonhead Fruit Mix         0.511   36.01763
Chiclets                           0.325   24.52499
Dots                               0.511   42.27208
Dum Dums                           0.034   39.46056
Fruit Chews                        0.034   43.08892
Fun Dip                            0.325   39.18550
Gobstopper                         0.453   46.78335
Haribo Gold Bears                  0.465   57.11974
Haribo Sour Bears                  0.465   51.41243
Haribo Twin Snakes                 0.465   42.17877
Jawbusters                         0.511   28.12744
Laffy Taffy                        0.116   41.38956
Lemonhead                          0.104   39.14106
Lifesavers big ring gummies        0.279   52.91139
Mike & Ike                         0.325   46.41172
Nerds                              0.325   55.35405
Nik L Nip                          0.976   22.44534
Now & Later                        0.325   39.44680
Pop Rocks                          0.837   41.26551
Red vines                          0.116   37.34852
Ring pop                           0.965   35.29076
Runts                              0.279   42.84914
Skittles original                  0.220   63.08514
Skittles wildberry                 0.220   55.10370
Smarties candy                     0.116   45.99583
Sour Patch Kids                    0.116   59.86400
Sour Patch Tricksters              0.116   52.82595
Starburst                          0.220   67.03763
Strawberry bon bons                0.058   34.57899
Super Bubble                       0.116   27.30386
Swedish Fish                       0.755   54.86111
Tootsie Pop                        0.325   48.98265
Trolli Sour Bites                  0.255   47.17323
Twizzlers                          0.116   45.46628
Warheads                           0.116   39.01190
Welch's Fruit Snacks               0.313   44.37552
```

```r
fruit.win <- candy[fruit.inds,]$winpercent
fruit.win
```

```
 [1] 52.34146 34.51768 36.01763 24.52499 42.27208 39.46056 43.08892 39.18550
 [9] 46.78335 57.11974 51.41243 42.17877 28.12744 41.38956 39.14106 52.91139
[17] 46.41172 55.35405 22.44534 39.44680 41.26551 37.34852 35.29076 42.84914
[25] 63.08514 55.10370 45.99583 59.86400 52.82595 67.03763 34.57899 27.30386
[33] 54.86111 48.98265 47.17323 45.46628 39.01190 44.37552
```

```r
fruit.mean<- mean(fruit.win)
fruit.mean
```

```
[1] 44.11974
```

It appears that the mean of fruity candy is much lower than the mean of chocolate candy, comparing 61% to 44%.

```r
choc.mean
```

```
[1] 60.92153
```

```r
fruit.mean
```

```
[1] 44.11974
```

> Q12. Is this difference statistically significant?

Yes, this t-test states that the values are not equal and the different is statistically different based on the p value.

```r
t.test(choc.win, fruit.win)
```

```

	Welch Two Sample t-test

data:  choc.win and fruit.win
t = 6.2582, df = 68.882, p-value = 2.871e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 11.44563 22.15795
sample estimates:
mean of x mean of y 
 60.92153  44.11974 
```

Overall Candy Rankings

Q13. What are the five least liked candy types in this set?

Nik L Nip, Boston Baked Beans, Chiclets, Super Bubble, Jawbusters

```
# Not that useful, sorts values numerically
sort(candy$winpercent)
```

```
 [1] 22.44534 23.41782 24.52499 27.30386 28.12744 29.70369 32.23100 32.26109
 [9] 33.43755 34.15896 34.51768 34.57899 34.72200 35.29076 36.01763 37.34852
[17] 37.72234 37.88719 38.01096 38.97504 39.01190 39.14106 39.18550 39.44680
[25] 39.46056 41.26551 41.38956 41.90431 42.17877 42.27208 42.84914 43.06890
[33] 43.08892 44.37552 45.46628 45.73675 45.99583 46.11650 46.29660 46.41172
[41] 46.78335 47.17323 47.82975 48.98265 49.52411 49.65350 50.34755 51.41243
[49] 52.34146 52.82595 52.91139 54.52645 54.86111 55.06407 55.10370 55.35405
[57] 55.37545 56.49050 56.91455 57.11974 57.21925 59.23612 59.52925 59.86400
[65] 60.80070 62.28448 63.08514 64.35334 65.71629 66.47068 66.57458 66.97173
[73] 67.03763 67.60294 69.48379 70.73564 71.46505 72.88790 73.09956 73.43499
[81] 76.67378 76.76860 81.64291 81.86626 84.18029
```

```
x<- c(10,1,100)
sort (x)
```

```
[1]   1  10 100
```

```
x[order(x)]
```

```
[1]   1  10 100
```

The **order()** function tells us how to arrange the elements of the input to make them sorted - i.e. how to order them.

We can determine the order of winprecent to make them sorted and use that order to arrange the whole dataset.

```
ord.inds <- order(candy$winpercent)
#tells which number candy is lowest to highest
head(candy[ord.inds, ])
```

```
                   chocolate fruity caramel peanutyalmondy nougat
Nik L Nip                  0      1       0              0      0
Boston Baked Beans         0      0       0              1      0
Chiclets                   0      1       0              0      0
Super Bubble               0      1       0              0      0
Jawbusters                 0      1       0              0      0
Root Beer Barrels          0      0       0              0      0
                   crispedricewafer hard bar pluribus sugarpercent pricepercent
Nik L Nip                         0    0   0        1        0.197        0.976
Boston Baked Beans                0    0   0        1        0.313        0.511
Chiclets                          0    0   0        1        0.046        0.325
Super Bubble                      0    0   0        0        0.162        0.116
Jawbusters                        0    1   0        1        0.093        0.511
Root Beer Barrels                 0    1   0        1        0.732        0.069
                   winpercent
Nik L Nip            22.44534
Boston Baked Beans   23.41782
Chiclets             24.52499
Super Bubble         27.30386
Jawbusters           28.12744
Root Beer Barrels    29.70369
```

Q14. What are the top 5 all time favorite candy types out of this set?

Reese's pieces, Snickers, Kit Kat, Twix, Reese's miniatures

```r
tail(candy[ord.inds,])
```

```
                          chocolate fruity caramel peanutyalmondy nougat
Reese's pieces                    1      0       0              1      0
Snickers                          1      0       1              1      1
Kit Kat                           1      0       0              0      0
Twix                              1      0       1              0      0
Reese's Miniatures                1      0       0              1      0
Reese's Peanut Butter cup         1      0       0              1      0
                          crispedricewafer hard bar pluribus sugarpercent
Reese's pieces                           0    0   0        1        0.406
Snickers                                 0    0   1        0        0.546
Kit Kat                                  1    0   1        0        0.313
Twix                                     1    0   1        0        0.546
Reese's Miniatures                       0    0   0        0        0.034
Reese's Peanut Butter cup                0    0   0        0        0.720
                          pricepercent winpercent
```

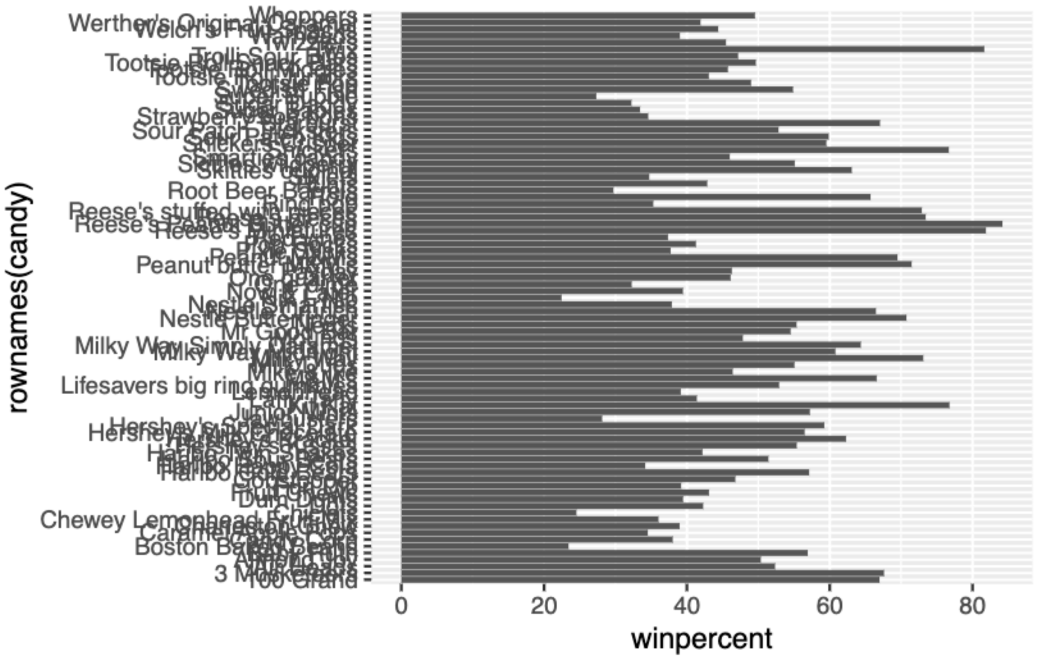
```
Reese's pieces                  0.651   73.43499
Snickers                        0.651   76.67378
Kit Kat                         0.511   76.76860
Twix                            0.906   81.64291
Reese's Miniatures              0.279   81.86626
Reese's Peanut Butter cup       0.651   84.18029
```

Another option

```
# order.inds <- order(cnady$winpercent, decreasing =T)
# pull heading
```

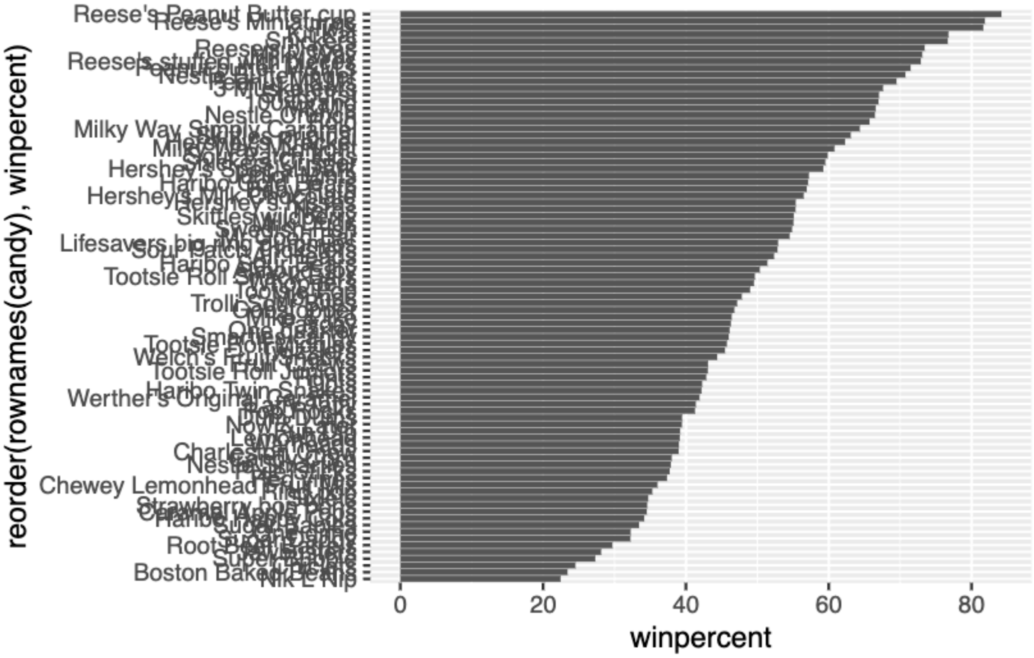
> Q15. Make a first barplot of candy ranking based on winpercent values.

```
ggplot (candy) +
  aes(winpercent, rownames(candy)) +
  geom_col()
```



> Q16. This is quite ugly, use the reorder() function to get the bars sorted by winpercent?

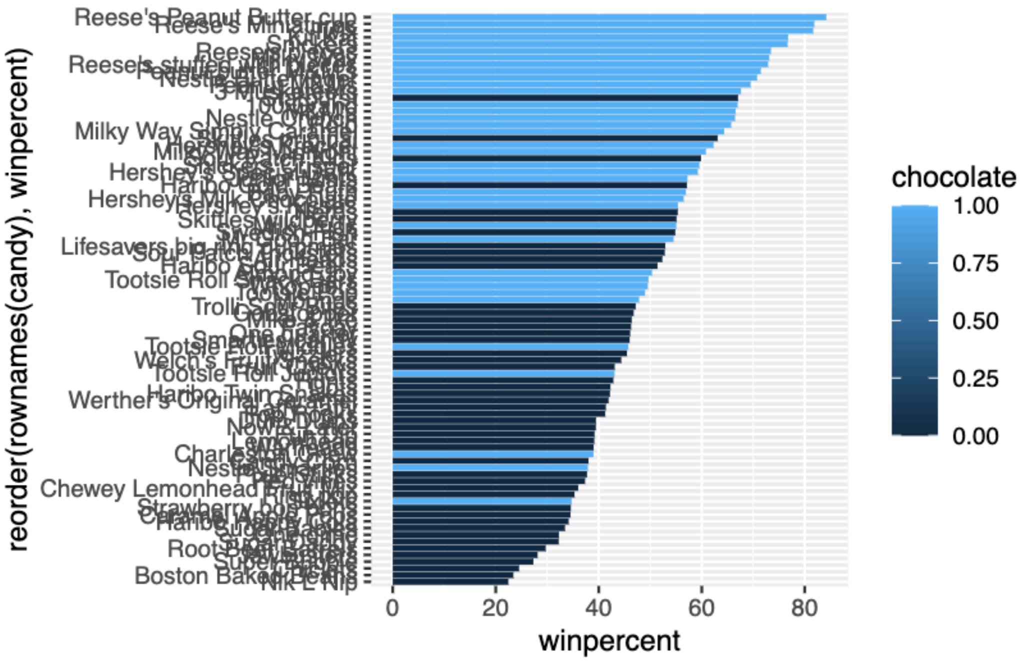
```
ggplot (candy) +
  aes(winpercent, reorder(rownames(candy), winpercent)) +
  geom_col()
```



Let's add some useful colors

This is not helpful, there isn't a scale from 0 to 1

```
ggplot (candy) +
  aes(winpercent, reorder(rownames(candy), winpercent), fill=chocolate) +
  geom_col()
```



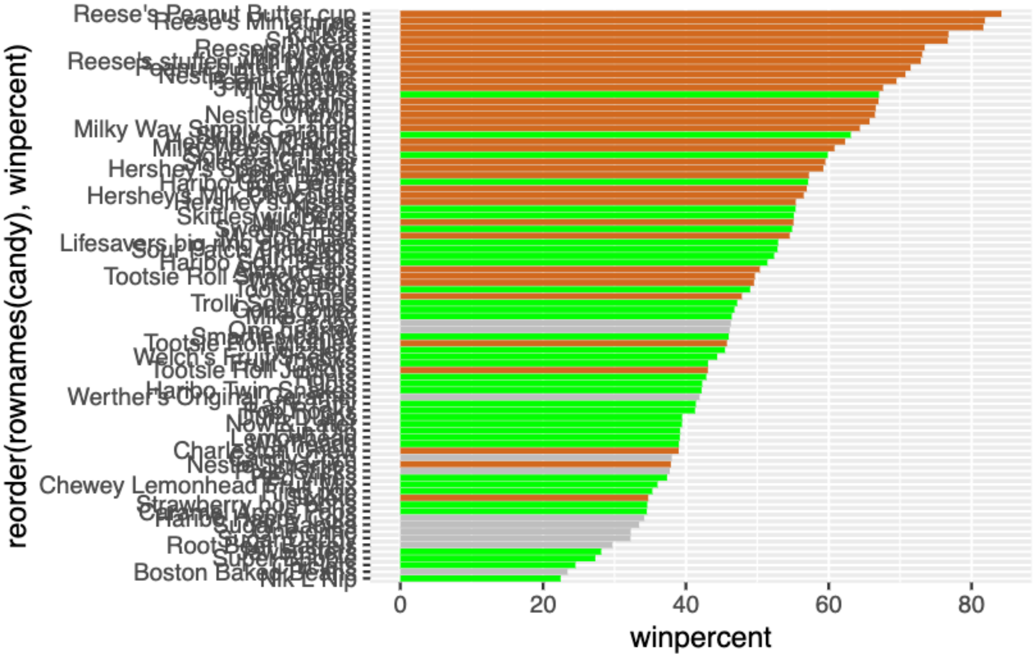
We need to make our own spearate color vector whwere we can spell our exactly what candy is colored a particular color.

```r
mycols <- rep("gray", nrow(candy))

mycols[candy$chocolate == 1] <- "chocolate"
mycols[candy$fruity == 1] <- "green"

mycols
```

```
 [1] "chocolate" "chocolate" "gray"      "gray"      "green"     "chocolate"
 [7] "chocolate" "gray"      "gray"      "green"     "chocolate" "green"
[13] "green"     "green"     "green"     "green"     "green"     "green"
[19] "green"     "gray"      "green"     "green"     "chocolate" "chocolate"
[25] "chocolate" "chocolate" "green"     "chocolate" "chocolate" "green"
[31] "green"     "green"     "chocolate" "chocolate" "green"     "chocolate"
[37] "chocolate" "chocolate" "chocolate" "chocolate" "chocolate" "green"
[43] "chocolate" "chocolate" "green"     "green"     "gray"      "chocolate"
[49] "gray"      "green"     "green"     "chocolate" "chocolate" "chocolate"
[55] "chocolate" "green"     "chocolate" "gray"      "green"     "chocolate"
[61] "green"     "green"     "chocolate" "green"     "chocolate" "chocolate"
[67] "green"     "green"     "green"     "green"     "gray"      "gray"
[73] "green"     "green"     "green"     "chocolate" "chocolate" "chocolate"
```

```
[79] "green"     "chocolate" "green"     "green"     "green"     "gray"
[85] "chocolate"
```

```
ggplot (candy) +
  aes(winpercent, reorder(rownames(candy), winpercent)) +
  geom_col(fill=mycols)
```



Both these functions work the same way to get out T/F values

```
c(1,0,1) ==1
```

```
[1]  TRUE FALSE  TRUE
```

```
as.logical(c(1,0,1))
```

```
[1]  TRUE FALSE  TRUE
```

Q17. What is the worst ranked chocolate candy?

Sixlets

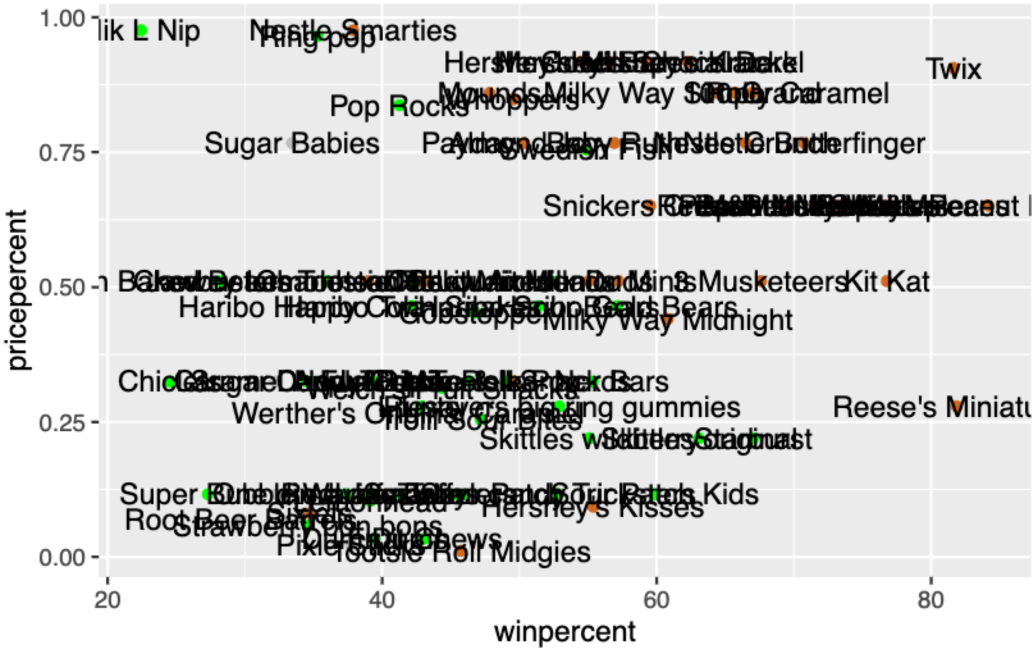
Q18. What is the best ranked fruity candy?

Starburst

Let's look at pricepercent

Making a plot of winpercent(x) versus pricepercent(y)

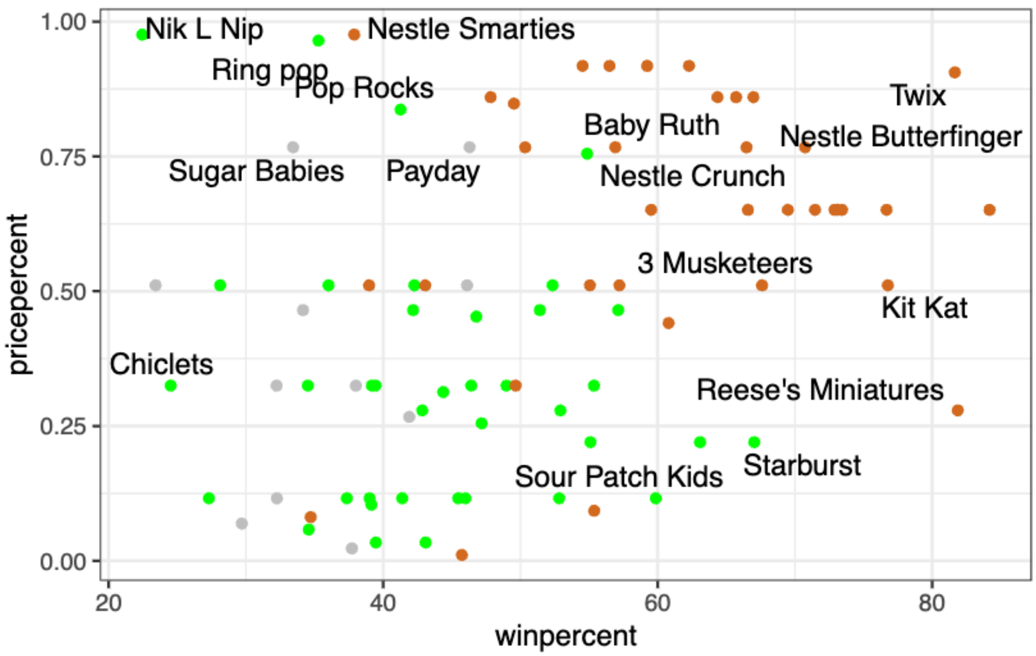
```
ggplot(candy) +
  aes(winpercent, pricepercent, label=rownames(candy)) +
  geom_point(col=mycols) +
  geom_text()
```



To avoid overplotting of the text labels, we can use an add-on package ggrepel

```
library(ggrepel)

ggplot(candy) +
  aes(winpercent, pricepercent, label=rownames(candy)) +
  geom_point(col=mycols) +
  geom_text_repel(max.overlaps = 6) +
  theme_bw()
```

```
Warning: ggrepel: 69 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```



> Q19. Which candy type is the highest ranked in terms of winpercent for the least money - i.e. offers the most bang for your buck?

Reese's miniatures

> Q20. What are the top 5 most expensive candy types in the dataset and of these which is the least popular?

Hershey's Krackel, Nik L Nip, Ring Pop, Nestle Smarties, and Hershey's Milk Chocolate. Nik L Nip is the least popular.

```
ord <- order(candy$pricepercent, decreasing = TRUE)
head( candy[ord,c(11,12)], n=5 )
```

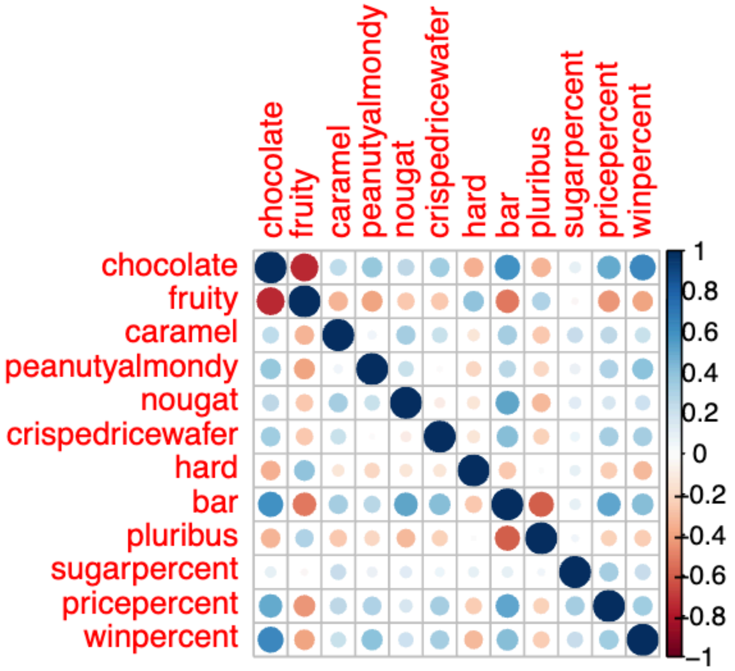
```
                         pricepercent winpercent
Nik L Nip                       0.976   22.44534
Nestle Smarties                 0.976   37.88719
Ring pop                        0.965   35.29076
Hershey's Krackel               0.918   62.28448
Hershey's Milk Chocolate        0.918   56.49050
```

Exploring the correlation structure

```
library(corrplot)
```

```
corrplot 0.95 loaded
```

```
cij <- cor(candy)
corrplot(cij)
```



Down the diagonal there is a perfect 1 correlation because its comparing to itself. -1 is high negative correlation, +1 is high positive correlation.

> Q22. Examining this plot what two variables are anti-correlated (i.e. have minus values)?

Fruity and chocolate.

> Q23. Similarly, what two variables are most positively correlated?

Winpercent and chocolate.

PCA

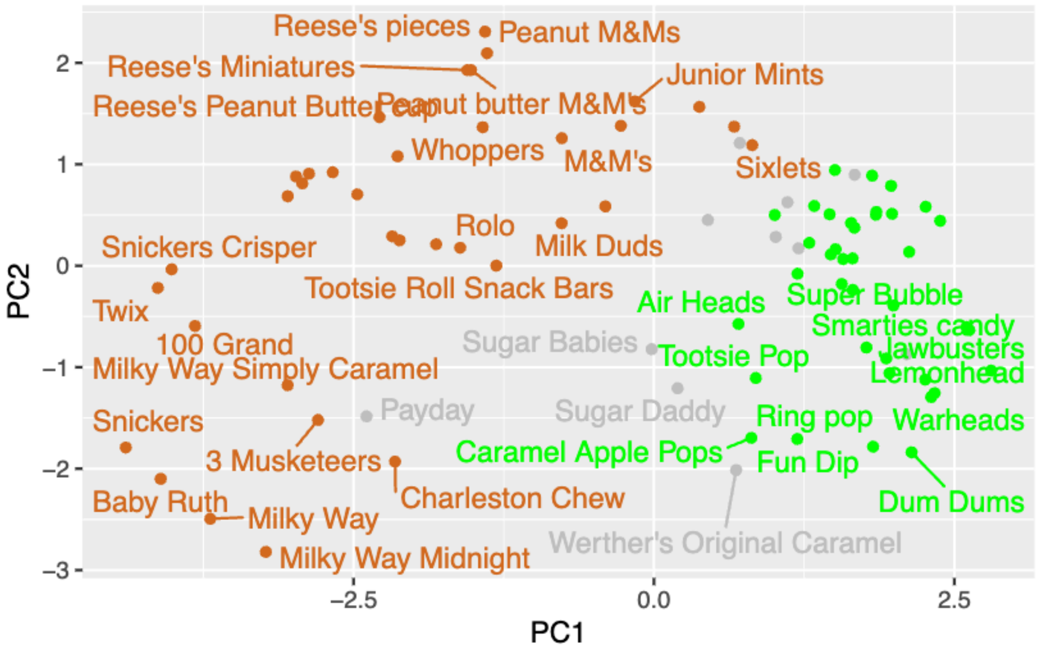
```
pca<- prcomp(candy, scale=T)
summary(pca)
```

```
Importance of components:
                          PC1    PC2    PC3     PC4    PC5     PC6     PC7
Standard deviation     2.0788 1.1378 1.1092 1.07533 0.9518 0.81923 0.81530
Proportion of Variance 0.3601 0.1079 0.1025 0.09636 0.0755 0.05593 0.05539
Cumulative Proportion  0.3601 0.4680 0.5705 0.66688 0.7424 0.79830 0.85369
                           PC8     PC9    PC10    PC11    PC12
Standard deviation     0.74530 0.67824 0.62349 0.43974 0.39760
Proportion of Variance 0.04629 0.03833 0.03239 0.01611 0.01317
Cumulative Proportion  0.89998 0.93832 0.97071 0.98683 1.00000
```

Let's plot our main resulst as our PCA "score plot"

```
ggplot(pca$x) +
  aes(PC1, PC2, label=rownames(pca$x))+
  geom_point(col=mycols) +
  geom_text_repel(col=mycols)
```

```
Warning: ggrepel: 48 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```



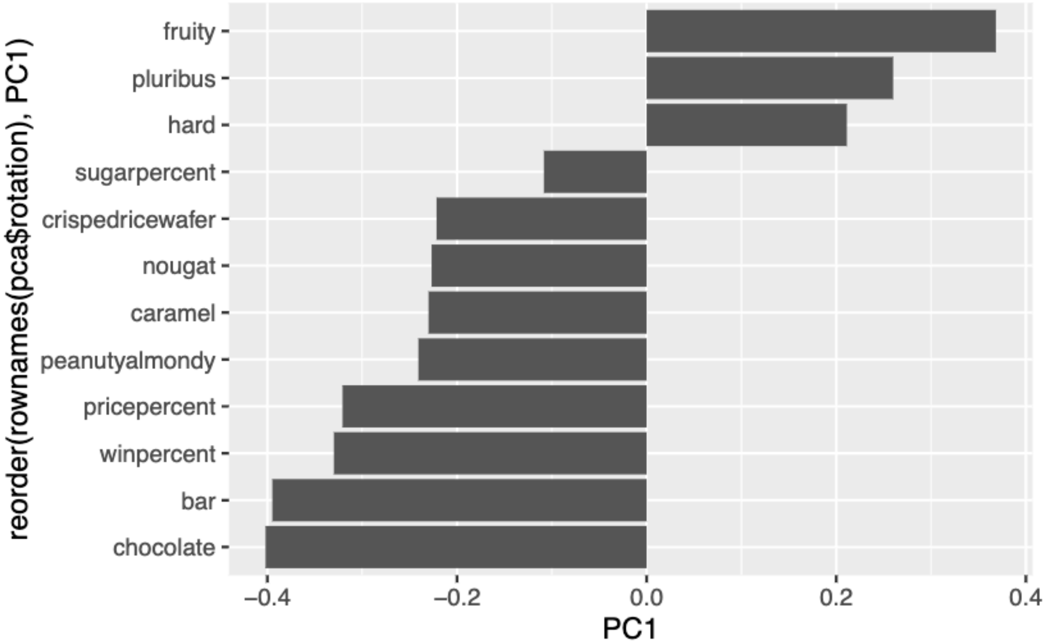
```r
summary(pca)
```

```
Importance of components:
                          PC1    PC2    PC3     PC4    PC5     PC6     PC7
Standard deviation     2.0788 1.1378 1.1092 1.07533 0.9518 0.81923 0.81530
Proportion of Variance 0.3601 0.1079 0.1025 0.09636 0.0755 0.05593 0.05539
Cumulative Proportion  0.3601 0.4680 0.5705 0.66688 0.7424 0.79830 0.85369
                           PC8     PC9    PC10    PC11    PC12
Standard deviation     0.74530 0.67824 0.62349 0.43974 0.39760
Proportion of Variance 0.04629 0.03833 0.03239 0.01611 0.01317
Cumulative Proportion  0.89998 0.93832 0.97071 0.98683 1.00000
```

> Q24. What original variables are picked up strongly by PC1 in the positive direction? Do these make sense to you?

PC1 has a strong positive correlation with our fruity candies, things that are hard and come with multiple items in one package. This makes sense because all these variables are connection, in that fruity candies are more often hard and come with multiple items in a package.

```r
#barplot(pca$rotation[,1], las=2, ylab="PC1 Contribution")

ggplot(pca$rotation) +
  aes(PC1, reorder(rownames(pca$rotation), PC1)) +
  geom_col()
```


reorder(rownames(pca$rotation), PC1)
fruity
pluribus
hard
sugarpercent
crispedricewafer
nougat
caramel
peanutyalmondy
pricepercent
winpercent
bar
chocolate
PC1