Class 09

Yaniv Iny (PID:A18090586)

Table of contents

Today we will examine data from 538 on common halloweedn candy. In particular we wil use ggplot,dplyr, and PCA to make sense of this multivariate dataset.

Importing candy data

```r
candy <- read.csv("candy-data.csv", 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?

```r
nrow(candy)
```

```
[1] 85
```

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

```r
sum(candy$fruit)
```

```
[1] 38
```

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

```r
candy["Mike & Ike", "winpercent"]
```

```
[1] 46.41172
```

How many chocolate candy are there in this dataset?

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

```
[1] 37
```

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

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

```
[1] 76.7686
```

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

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

```
[1] 49.6535
```

```r
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	

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

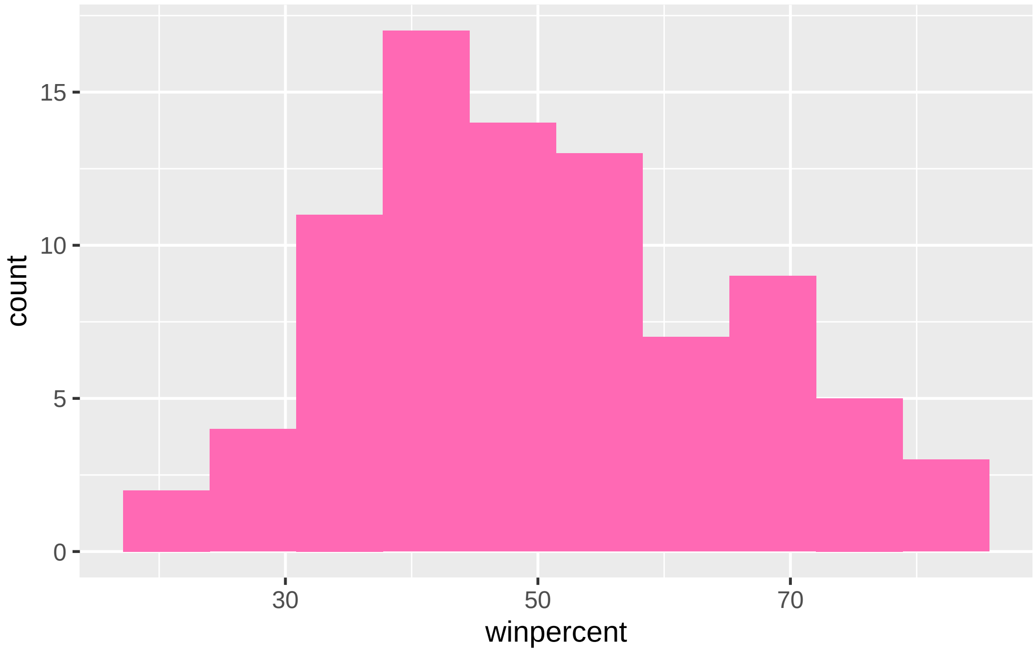
The scale of the winpercent variable seems to be on a differnt scale than any of the other data columns. It seems to be on a scale of (0-100%, rather than 0-1)

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

That it does or does not contain chocolate

> Q8. Plot a histogram of winpercent values

```r
library(ggplot2)

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



Q9. Is the distribution of winpercent values symmetrical?

No > Q10. Is the center of the distribution above or below 50%?

Center seems to be below 50% to validate this we can do

```r
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

```r
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 find their "winpercent" calues

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

```
                              chocolate fruity caramel peanutyalmondy nougat
Snickers Crisper                      1      0       1              1      0
Sour Patch Kids                       0      1       0              0      0
Pop Rocks                             0      1       0              0      0
```

```
Ring pop                              0      1       0              0      0
Milky Way Midnight                    1      0       1              0      1
Reese's stuffed with pieces           1      0       0              1      0
Skittles wildberry                    0      1       0              0      0
Ring pop.1                            0      1       0              0      0
Runts                                 0      1       0              0      0
Rolo                                  1      0       1              0      0
Tootsie Roll Juniors                  1      0       0              0      0
Sugar Babies                          0      0       1              0      0
Snickers Crisper.1                    1      0       1              1      0
Reese's stuffed with pieces.1         1      0       0              1      0
Super Bubble                          0      1       0              0      0
Sixlets                               1      0       0              0      0
Smarties candy                        0      1       0              0      0
Payday                                0      0       0              1      1
Reese's pieces                        1      0       0              1      0
Strawberry bon bons                   0      1       0              0      0
Snickers Crisper.2                    1      0       1              1      0
Starburst                             0      1       0              0      0
Twizzlers                             0      1       0              0      0
Werther's Original Caramel            0      0       1              0      0
Super Bubble.1                        0      1       0              0      0
Sugar Daddy                           0      0       1              0      0
Snickers                              1      0       1              1      1
M&M's                                 1      0       0              0      0
Milky Way                             1      0       1              0      1
Tootsie Roll Juniors.1                1      0       0              0      0
Runts.1                               0      1       0              0      0
Peanut M&Ms                           1      0       0              1      0
Nestle Butterfinger                   1      0       0              1      0
Nik L Nip                             0      1       0              0      0
Pixie Sticks                          0      0       0              0      0
Twizzlers.1                           0      1       0              0      0
Pixie Sticks.1                        0      0       0              0      0
                              crispedricewafer hard bar pluribus sugarpercent
Snickers Crisper                             1    0   1        0        0.604
Sour Patch Kids                              0    0   0        1        0.069
Pop Rocks                                    0    1   0        1        0.604
Ring pop                                     0    1   0        0        0.732
Milky Way Midnight                           0    0   1        0        0.732
Reese's stuffed with pieces                  0    0   0        0        0.988
Skittles wildberry                           0    0   0        1        0.941
Ring pop.1                                   0    1   0        0        0.732
```

```
Runts                                        0    1   0        1         0.872
Rolo                                         0    0   0        1         0.860
Tootsie Roll Juniors                         0    0   0        0         0.313
Sugar Babies                                 0    0   0        1         0.965
Snickers Crisper.1                           1    0   1        0         0.604
Reese's stuffed with pieces.1                0    0   0        0         0.988
Super Bubble                                 0    0   0        0         0.162
Sixlets                                      0    0   0        1         0.220
Smarties candy                               0    1   0        1         0.267
Payday                                       0    0   1        0         0.465
Reese's pieces                               0    0   0        1         0.406
Strawberry bon bons                          0    1   0        1         0.569
Snickers Crisper.2                           1    0   1        0         0.604
Starburst                                    0    0   0        1         0.151
Twizzlers                                    0    0   0        0         0.220
Werther's Original Caramel                   0    1   0        0         0.186
Super Bubble.1                               0    0   0        0         0.162
Sugar Daddy                                  0    0   0        0         0.418
Snickers                                     0    0   1        0         0.546
M&M's                                        0    0   0        1         0.825
Milky Way                                    0    0   1        0         0.604
Tootsie Roll Juniors.1                       0    0   0        0         0.313
Runts.1                                      0    1   0        1         0.872
Peanut M&Ms                                  0    0   0        1         0.593
Nestle Butterfinger                          0    0   1        0         0.604
Nik L Nip                                    0    0   0        1         0.197
Pixie Sticks                                 0    0   0        1         0.093
Twizzlers.1                                  0    0   0        0         0.220
Pixie Sticks.1                               0    0   0        1         0.093
                              pricepercent winpercent
Snickers Crisper                     0.651   59.52925
Sour Patch Kids                      0.116   59.86400
Pop Rocks                            0.837   41.26551
Ring pop                             0.965   35.29076
Milky Way Midnight                   0.441   60.80070
Reese's stuffed with pieces          0.651   72.88790
Skittles wildberry                   0.220   55.10370
Ring pop.1                           0.965   35.29076
Runts                                0.279   42.84914
Rolo                                 0.860   65.71629
Tootsie Roll Juniors                 0.511   43.06890
Sugar Babies                         0.767   33.43755
Snickers Crisper.1                   0.651   59.52925
```

```
Reese's stuffed with pieces.1        0.651   72.88790
Super Bubble                         0.116   27.30386
Sixlets                              0.081   34.72200
Smarties candy                       0.116   45.99583
Payday                               0.767   46.29660
Reese's pieces                       0.651   73.43499
Strawberry bon bons                  0.058   34.57899
Snickers Crisper.2                   0.651   59.52925
Starburst                            0.220   67.03763
Twizzlers                            0.116   45.46628
Werther's Original Caramel           0.267   41.90431
Super Bubble.1                       0.116   27.30386
Sugar Daddy                          0.325   32.23100
Snickers                             0.651   76.67378
M&M's                                0.651   66.57458
Milky Way                            0.651   73.09956
Tootsie Roll Juniors.1               0.511   43.06890
Runts.1                              0.279   42.84914
Peanut M&Ms                          0.651   69.48379
Nestle Butterfinger                  0.767   70.73564
Nik L Nip                            0.976   22.44534
Pixie Sticks                         0.023   37.72234
Twizzlers.1                          0.116   45.46628
Pixie Sticks.1                       0.023   37.72234
```

- Step 3 summarize these values

```
mean(choc.win)
```

```
[1] 60.92153
```

- Step 4 find all "fruit" candy

```
fruit.inds <- candy$fruit == 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
```

- Step 5 find their "winpercent" calues

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

```
                              chocolate fruity caramel peanutyalmondy nougat
Reese's Miniatures                    1      0       0              1      0
M&M's                                 1      0       0              0      0
Milk Duds                             1      0       1              0      0
Hershey's Krackel                     1      0       0              0      0
Nerds                                 0      1       0              0      0
Milky Way Simply Caramel              1      0       1              0      0
Nestle Butterfinger                   1      0       0              1      0
Milky Way Simply Caramel.1            1      0       1              0      0
Now & Later                           0      1       0              0      0
Rolo                                  1      0       1              0      0
```

```
Red vines                             0      1       0              0      0
Nerds.1                               0      1       0              0      0
Junior Mints                          1      0       0              0      0
Mr Good Bar                           1      0       0              1      0
Milky Way Simply Caramel.2            1      0       1              0      0
Reese's Miniatures.1                  1      0       0              1      0
Now & Later.1                         0      1       0              0      0
Reese's stuffed with pieces           1      0       0              1      0
Haribo Twin Snakes                    0      1       0              0      0
Milky Way Simply Caramel.3            1      0       1              0      0
Mr Good Bar.1                         1      0       0              1      0
Milky Way                             1      0       1              0      1
Mike & Ike                            0      1       0              0      0
Nerds.2                               0      1       0              0      0
Nestle Smarties                       1      0       0              0      0
Reese's stuffed with pieces.1         1      0       0              1      0
Nik L Nip                             0      1       0              0      0
Runts                                 0      1       0              0      0
Reese's Miniatures.2                  1      0       0              1      0
Sour Patch Kids                       0      1       0              0      0
M&M's.1                               1      0       0              0      0
Jawbusters                            0      1       0              0      0
Reese's pieces                        1      0       0              1      0
Peanut M&Ms                           1      0       0              1      0
Payday                                0      0       0              1      1
Nik L Nip.1                           0      1       0              0      0
Milky Way Simply Caramel.4            1      0       1              0      0
Nestle Crunch                         1      0       0              0      0
                              crispedricewafer hard bar pluribus sugarpercent
Reese's Miniatures                           0    0   0        0        0.034
M&M's                                        0    0   0        1        0.825
Milk Duds                                    0    0   0        1        0.302
Hershey's Krackel                            1    0   1        0        0.430
Nerds                                        0    1   0        1        0.848
Milky Way Simply Caramel                     0    0   1        0        0.965
Nestle Butterfinger                          0    0   1        0        0.604
Milky Way Simply Caramel.1                   0    0   1        0        0.965
Now & Later                                  0    0   0        1        0.220
Rolo                                         0    0   0        1        0.860
Red vines                                    0    0   0        1        0.581
Nerds.1                                      0    1   0        1        0.848
Junior Mints                                 0    0   0        1        0.197
Mr Good Bar                                  0    0   1        0        0.313
```

```
Milky Way Simply Caramel.2                   0    0   1        0        0.965
Reese's Miniatures.1                         0    0   0        0        0.034
Now & Later.1                                0    0   0        1        0.220
Reese's stuffed with pieces                  0    0   0        0        0.988
Haribo Twin Snakes                           0    0   0        1        0.465
Milky Way Simply Caramel.3                   0    0   1        0        0.965
Mr Good Bar.1                                0    0   1        0        0.313
Milky Way                                    0    0   1        0        0.604
Mike & Ike                                   0    0   0        1        0.872
Nerds.2                                      0    1   0        1        0.848
Nestle Smarties                              0    0   0        1        0.267
Reese's stuffed with pieces.1                0    0   0        0        0.988
Nik L Nip                                    0    0   0        1        0.197
Runts                                        0    1   0        1        0.872
Reese's Miniatures.2                         0    0   0        0        0.034
Sour Patch Kids                              0    0   0        1        0.069
M&M's.1                                      0    0   0        1        0.825
Jawbusters                                   0    1   0        1        0.093
Reese's pieces                               0    0   0        1        0.406
Peanut M&Ms                                  0    0   0        1        0.593
Payday                                       0    0   1        0        0.465
Nik L Nip.1                                  0    0   0        1        0.197
Milky Way Simply Caramel.4                   0    0   1        0        0.965
Nestle Crunch                                1    0   1        0        0.313
                              pricepercent winpercent
Reese's Miniatures                   0.279   81.86626
M&M's                                0.651   66.57458
Milk Duds                            0.511   55.06407
Hershey's Krackel                    0.918   62.28448
Nerds                                0.325   55.35405
Milky Way Simply Caramel             0.860   64.35334
Nestle Butterfinger                  0.767   70.73564
Milky Way Simply Caramel.1           0.860   64.35334
Now & Later                          0.325   39.44680
Rolo                                 0.860   65.71629
Red vines                            0.116   37.34852
Nerds.1                              0.325   55.35405
Junior Mints                         0.511   57.21925
Mr Good Bar                          0.918   54.52645
Milky Way Simply Caramel.2           0.860   64.35334
Reese's Miniatures.1                 0.279   81.86626
Now & Later.1                        0.325   39.44680
Reese's stuffed with pieces          0.651   72.88790
```

```
Haribo Twin Snakes                    0.465   42.17877
Milky Way Simply Caramel.3            0.860   64.35334
Mr Good Bar.1                         0.918   54.52645
Milky Way                             0.651   73.09956
Mike & Ike                            0.325   46.41172
Nerds.2                               0.325   55.35405
Nestle Smarties                       0.976   37.88719
Reese's stuffed with pieces.1         0.651   72.88790
Nik L Nip                             0.976   22.44534
Runts                                 0.279   42.84914
Reese's Miniatures.2                  0.279   81.86626
Sour Patch Kids                       0.116   59.86400
M&M's.1                               0.651   66.57458
Jawbusters                            0.511   28.12744
Reese's pieces                        0.651   73.43499
Peanut M&Ms                           0.651   69.48379
Payday                                0.767   46.29660
Nik L Nip.1                           0.976   22.44534
Milky Way Simply Caramel.4            0.860   64.35334
Nestle Crunch                         0.767   66.47068
```

- Step 6 summarize these values

```r
fruit <- read.csv("candy-data.csv")
fruit$win <- as.numeric(as.character(fruit$win))
mean(fruit.win, na.rm=TRUE)
```

```
[1] 44.11974
```

- Step 7 compare the two summary values 60 vs 40

  Q12. Is this difference statistically significant?

YES

```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 
```

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

Nik L Nip, Boston Baked beans, Chiclets, superbubble, jawbreaker

```r
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
```

```r
ord.inds <- order(candy$winpercent, decreasing = F)
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
```

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

```
[1] 2 1 3
```

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

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

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

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

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

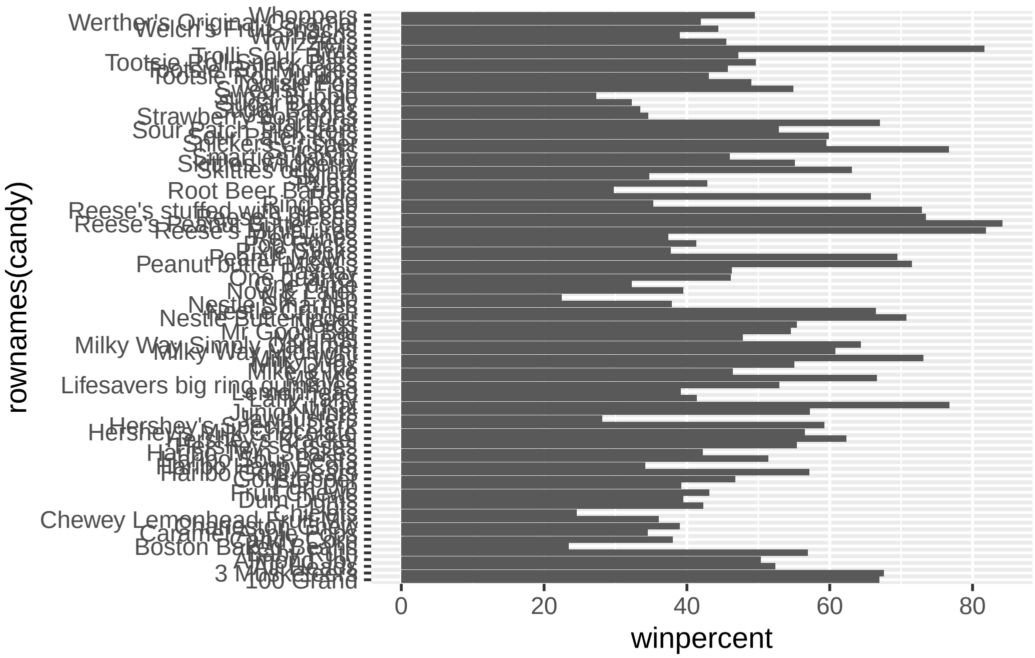
```
ord.inds <- order(candy$winpercent, decreasing = T)
head( candy[ord.inds, ])
```

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

```
Reese's pieces                          0    0   0         1         0.406
                          pricepercent winpercent
Reese's Peanut Butter cup        0.651   84.18029
Reese's Miniatures               0.279   81.86626
Twix                             0.906   81.64291
Kit Kat                          0.511   76.76860
Snickers                         0.651   76.67378
Reese's pieces                   0.651   73.43499
```

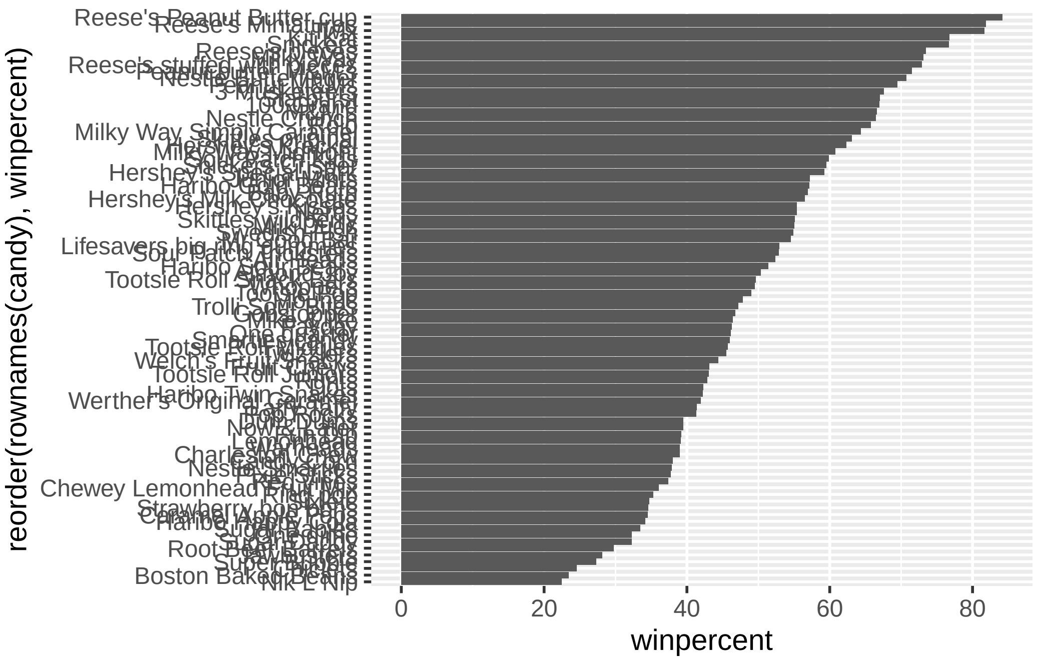
Q15. Make a first barplot of candy ranking based on winpercent values.

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



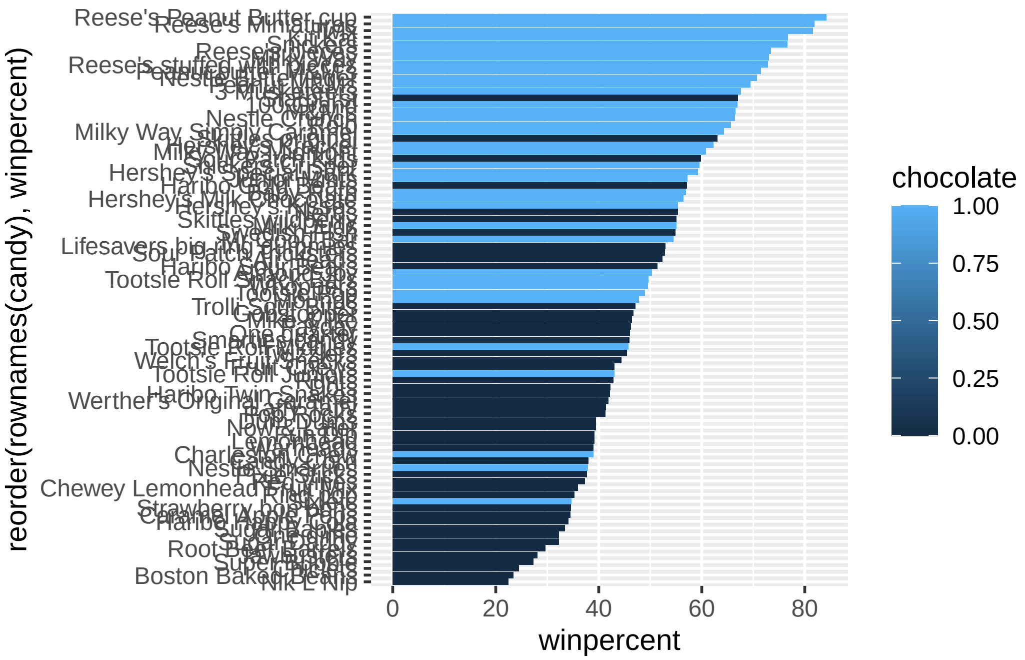
Q16. This is quite ugly, use the reorder() function to get the bars sorted by winpercent?

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



Time to add some useful color

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



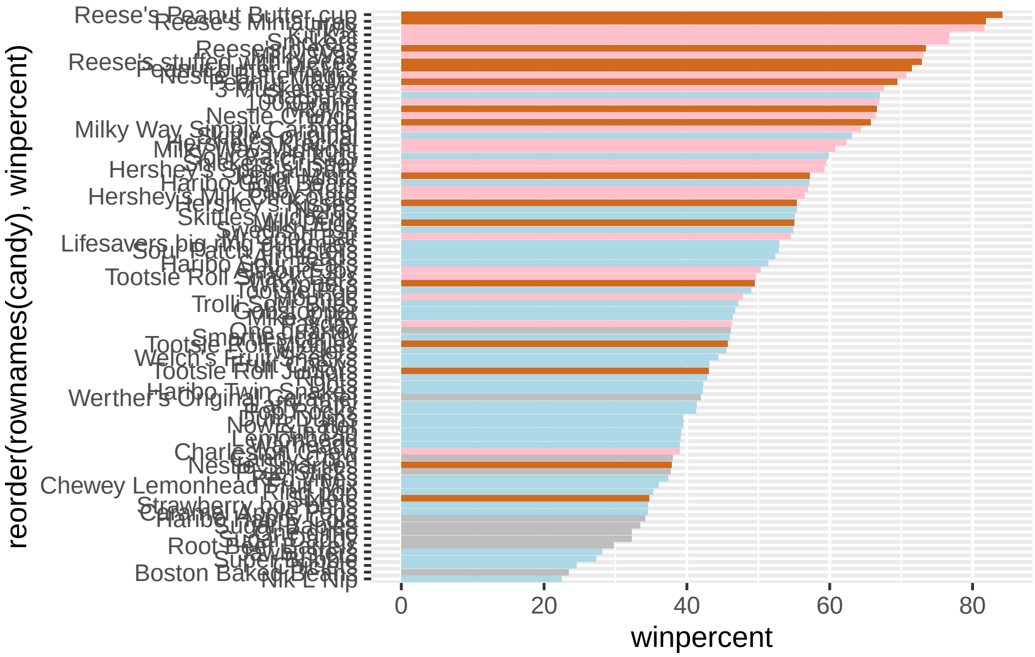
We need to make our own seperate color vector where we can spell out exactly what candy is colored a particular color.

```
mycols <- rep("gray",nrow(candy))
mycols[candy$chocolate==1] <- "chocolate"
mycols[candy$fruity==1] <- "lightblue"
mycols[candy$bar==1] <- "pink"
mycols
```

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

```
[85] "chocolate"
```

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



Q17. What is the worst ranked chocolate candy?

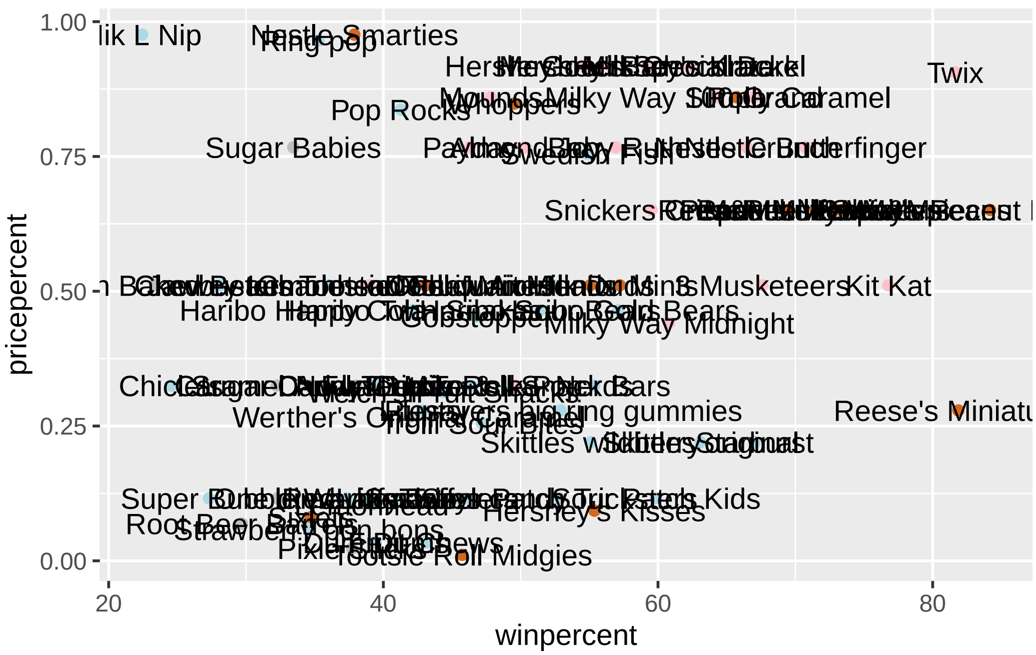
Boston Baked Beans > Q18. What is the best ranked fruity candy?

Starbust

Taking a look at pricepercent

Make a plot of winpercent (x-axis) vs pricepercent (y-axis)

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

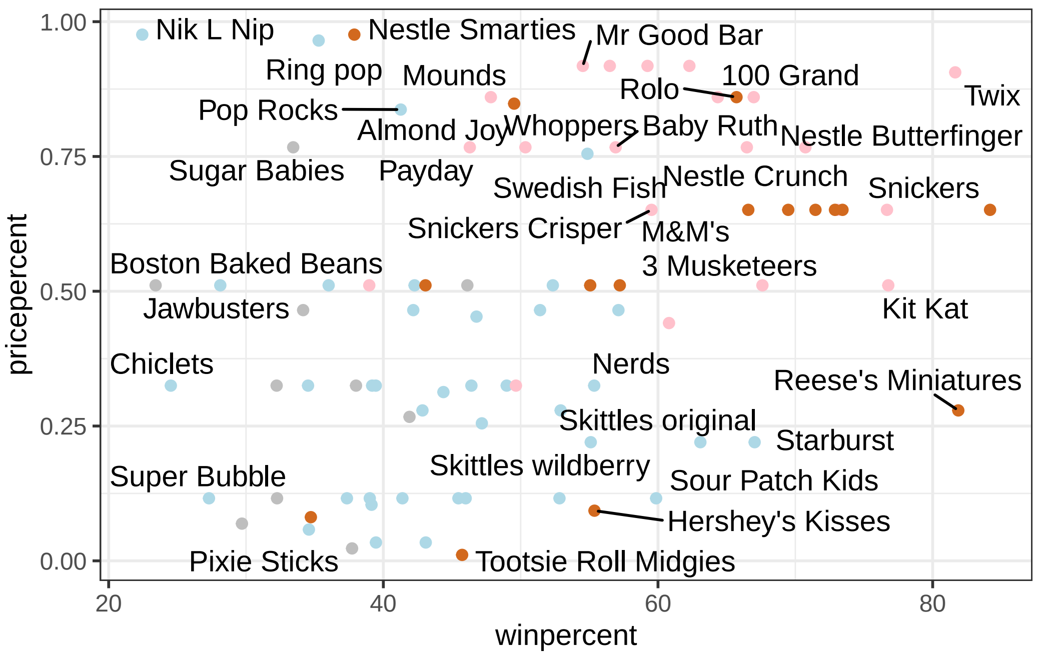


To avoid the overplotting of the text labels we can use the add on package **ggrepel**

```
library(ggrepel)

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

```
Warning in geom_text_repel(maxoverlaps = 6): Ignoring unknown parameters:
`maxoverlaps`
```

```
Warning: ggrepel: 50 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?

Reeses Miniatures

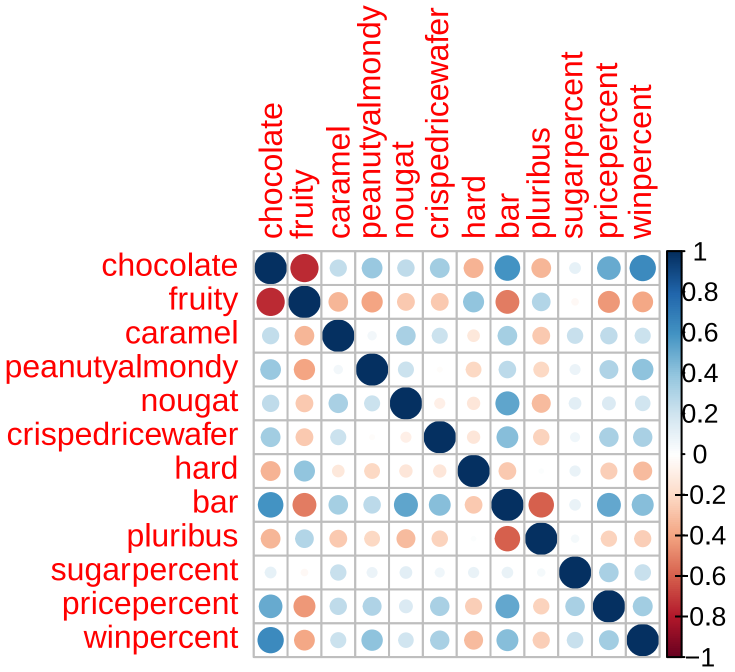
> Q20. What are the top 5 most expensive candy types in the dataset and of these which is the least popular?

Twix, Hershey, NikLnip, Ringpop, Nestle Smarties, The least populat is the Liklnip ## Exploring the correlation structure

```
library(corrplot)
```

```
corrplot 0.95 loaded
```

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



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

Two values that are anti correlated is chocolate and fruit candies.

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

Most positively correlated is the winpercent to the chocolate

Principal Component Analysis

Let's apply PCA using the prcom() function to our candy dataset remembering to set the scale=TRUE argument.

```r
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
```

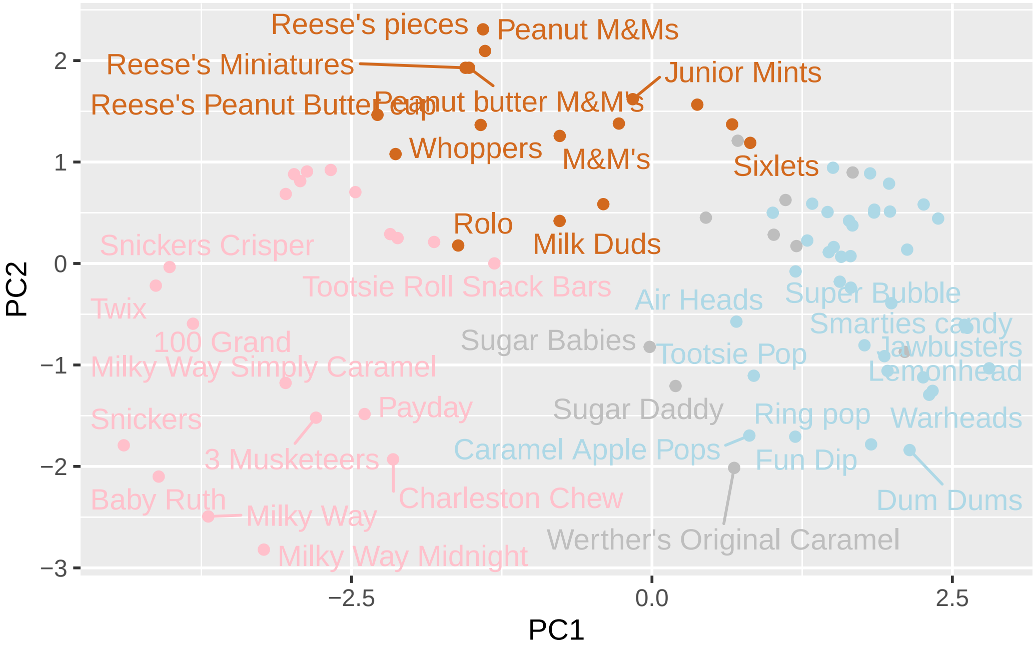
```
attributes(pca)
```

```
$names
[1] "sdev"     "rotation" "center"   "scale"    "x"

$class
[1] "prcomp"
```

Lets plot our mian results as our PCA “score plot”

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

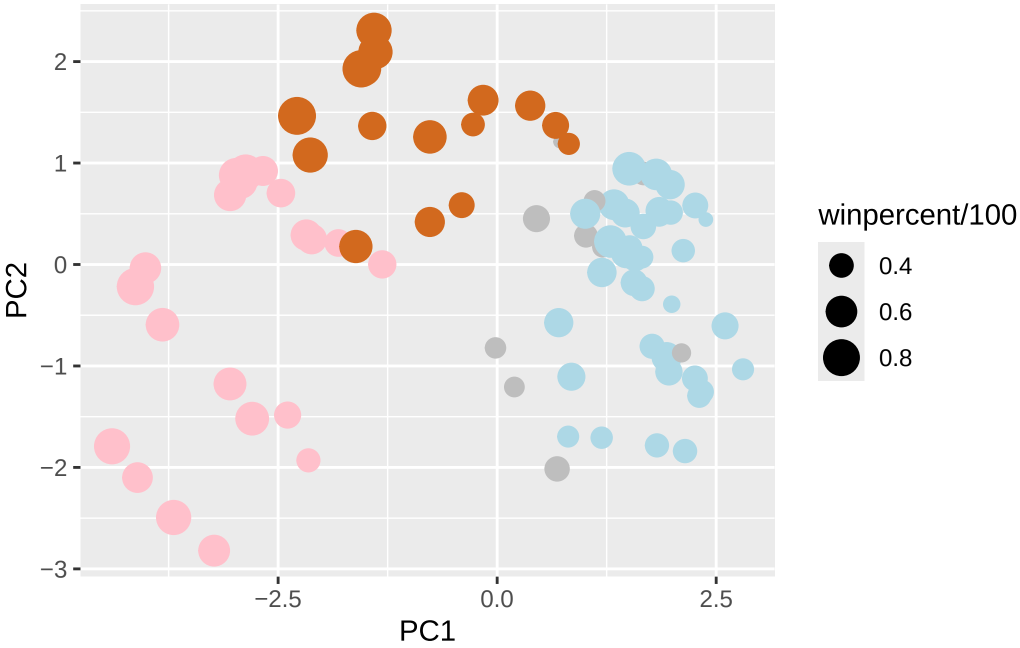
```
Warning: ggrepel: 48 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```



Make a new data-frame with our PCA results and candy data

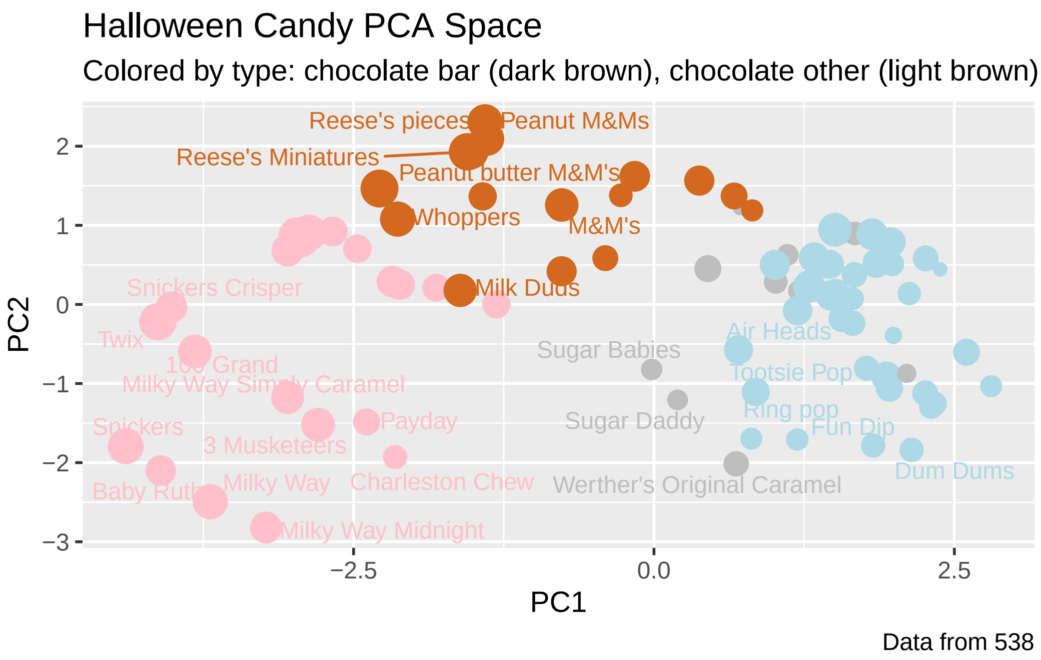
```
my_data <- cbind(candy, pca$x[,1:3])
```

```
p <- ggplot(my_data) +
        aes(x=PC1, y=PC2,
            size=winpercent/100,
            text=rownames(my_data),
            label=rownames(my_data)) +
        geom_point(col=mycols)

p
```



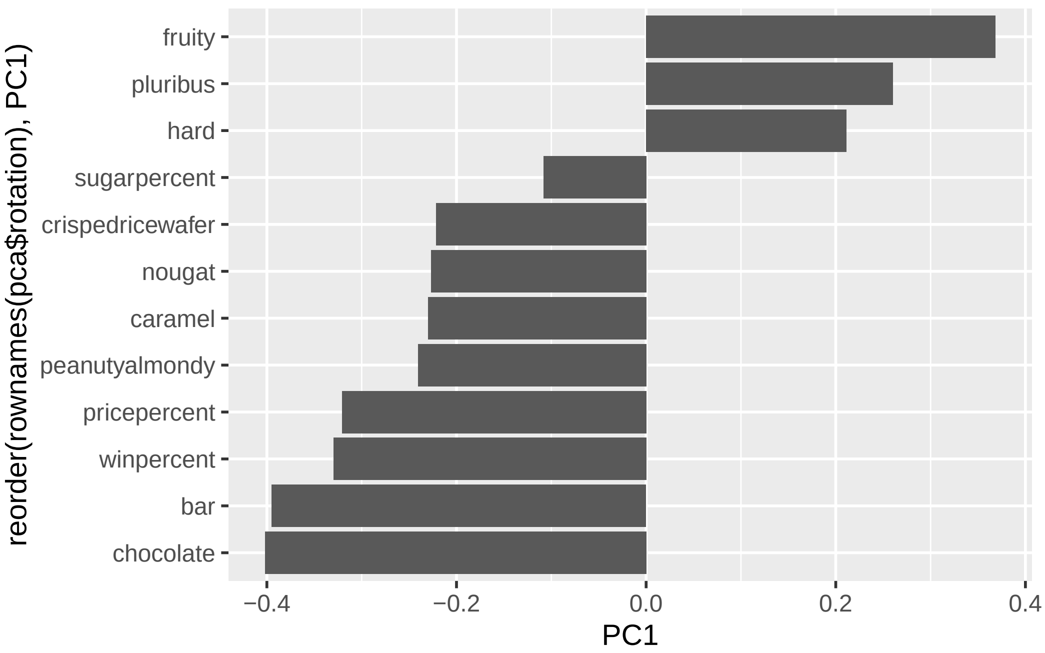
```
library(ggrepel)

p + geom_text_repel(size=3.3, col=mycols, max.overlaps = 7)  +
  theme(legend.position = "none") +
  labs(title="Halloween Candy PCA Space",
       subtitle="Colored by type: chocolate bar (dark brown), chocolate other (light brown),
       caption="Data from 538")
```

```
Warning: ggrepel: 59 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```



Finally letsl ook at how the original variables contribute to the pCs, start with PC1

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



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

Fruity, Pluribus, and Hard, yes this does make sense as most fruity candy is hard and has multiple pieces in the package.