Class 10: Halloween Mini-Project

Juan Gonzalez (PID: A69036681)

#Importing candy data

```r
candy_file <- "candy-data.csv"

candy = read.csv(candy_file, 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
```

Question1: How many different candy types are in this dataset? 85 candy types

```r
nrow(candy)
```

```
[1] 85
```

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

```r
sum(candy$fruity==1)
```

```
[1] 38
```

#What is your favorite candy?

One of the most interesting variables in the dataset is `winpercent`. For a given candy this value is the percentage of people who prefer this candy over another randomly chosen candy from the dataset (what 538 term a matchup). Higher values indicate a more popular candy.

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

```r
candy
```

```
                            chocolate fruity caramel peanutyalmondy nougat
100 Grand                           1      0       1              0      0
3 Musketeers                        1      0       0              0      1
One dime                            0      0       0              0      0
One quarter                         0      0       0              0      0
Air Heads                           0      1       0              0      0
Almond Joy                          1      0       0              1      0
Baby Ruth                           1      0       1              1      1
Boston Baked Beans                  0      0       0              1      0
Candy Corn                          0      0       0              0      0
Caramel Apple Pops                  0      1       1              0      0
Charleston Chew                     1      0       0              0      1
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 Happy Cola                   0      0       0              0      0
Haribo Sour Bears                   0      1       0              0      0
Haribo Twin Snakes                  0      1       0              0      0
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
Jawbusters                          0      1       0              0      0
```

Junior Mints	1	0	0	0	0
Kit Kat	1	0	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
Peanut butter M&M's	1	0	0	1	0
M&M's	1	0	0	0	0
Mike & Ike	0	1	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
Nerds	0	1	0	0	0
Nestle Butterfinger	1	0	0	1	0
Nestle Crunch	1	0	0	0	0
Nik L Nip	0	1	0	0	0
Now & Later	0	1	0	0	0
Payday	0	0	0	1	1
Peanut M&Ms	1	0	0	1	0
Pixie Sticks	0	0	0	0	0
Pop Rocks	0	1	0	0	0
Red vines	0	1	0	0	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
Ring pop	0	1	0	0	0
Rolo	1	0	1	0	0
Root Beer Barrels	0	0	0	0	0
Runts	0	1	0	0	0
Sixlets	1	0	0	0	0
Skittles original	0	1	0	0	0
Skittles wildberry	0	1	0	0	0
Nestle Smarties	1	0	0	0	0
Smarties candy	0	1	0	0	0
Snickers	1	0	1	1	1
Snickers Crisper	1	0	1	1	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

```
Sugar Babies                        0      0       1              0      0
Sugar Daddy                         0      0       1              0      0
Super Bubble                        0      1       0              0      0
Swedish Fish                        0      1       0              0      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
Trolli Sour Bites                   0      1       0              0      0
Twix                                1      0       1              0      0
Twizzlers                           0      1       0              0      0
Warheads                            0      1       0              0      0
Welch's Fruit Snacks                0      1       0              0      0
Werther's Original Caramel          0      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
One dime                                   0    0   0        0        0.011
One quarter                                0    0   0        0        0.011
Air Heads                                  0    0   0        0        0.906
Almond Joy                                 0    0   1        0        0.465
Baby Ruth                                  0    0   1        0        0.604
Boston Baked Beans                         0    0   0        1        0.313
Candy Corn                                 0    0   0        1        0.906
Caramel Apple Pops                         0    0   0        0        0.604
Charleston Chew                            0    0   1        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 Happy Cola                          0    0   0        1        0.465
Haribo Sour Bears                          0    0   0        1        0.465
Haribo Twin Snakes                         0    0   0        1        0.465
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
Jawbusters                                 0    1   0        1        0.093
```

Junior Mints	0	0	0	1	0.197
Kit Kat	1	0	1	0	0.313
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
Peanut butter M&M's	0	0	0	1	0.825
M&M's	0	0	0	1	0.825
Mike & Ike	0	0	0	1	0.872
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
Nerds	0	1	0	1	0.848
Nestle Butterfinger	0	0	1	0	0.604
Nestle Crunch	1	0	1	0	0.313
Nik L Nip	0	0	0	1	0.197
Now & Later	0	0	0	1	0.220
Payday	0	0	1	0	0.465
Peanut M&Ms	0	0	0	1	0.593
Pixie Sticks	0	0	0	1	0.093
Pop Rocks	0	1	0	1	0.604
Red vines	0	0	0	1	0.581
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
Ring pop	0	1	0	0	0.732
Rolo	0	0	0	1	0.860
Root Beer Barrels	0	1	0	1	0.732
Runts	0	1	0	1	0.872
Sixlets	0	0	0	1	0.220
Skittles original	0	0	0	1	0.941
Skittles wildberry	0	0	0	1	0.941
Nestle Smarties	0	0	0	1	0.267
Smarties candy	0	1	0	1	0.267
Snickers	0	0	1	0	0.546
Snickers Crisper	1	0	1	0	0.604
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

```
Sugar Babies                               0    0   0        1        0.965
Sugar Daddy                                0    0   0        0        0.418
Super Bubble                               0    0   0        0        0.162
Swedish Fish                               0    0   0        1        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
Trolli Sour Bites                          0    0   0        1        0.313
Twix                                       1    0   1        0        0.546
Twizzlers                                  0    0   0        0        0.220
Warheads                                   0    1   0        0        0.093
Welch's Fruit Snacks                       0    0   0        1        0.313
Werther's Original Caramel                 0    1   0        0        0.186
Whoppers                                   1    0   0        1        0.872
                            pricepercent winpercent
100 Grand                          0.860   66.97173
3 Musketeers                       0.511   67.60294
One dime                           0.116   32.26109
One quarter                        0.511   46.11650
Air Heads                          0.511   52.34146
Almond Joy                         0.767   50.34755
Baby Ruth                          0.767   56.91455
Boston Baked Beans                 0.511   23.41782
Candy Corn                         0.325   38.01096
Caramel Apple Pops                 0.325   34.51768
Charleston Chew                    0.511   38.97504
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 Happy Cola                  0.465   34.15896
Haribo Sour Bears                  0.465   51.41243
Haribo Twin Snakes                 0.465   42.17877
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
Jawbusters                         0.511   28.12744
```

Junior Mints	0.511	57.21925
Kit Kat	0.511	76.76860
Laffy Taffy	0.116	41.38956
Lemonhead	0.104	39.14106
Lifesavers big ring gummies	0.279	52.91139
Peanut butter M&M's	0.651	71.46505
M&M's	0.651	66.57458
Mike & Ike	0.325	46.41172
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
Nerds	0.325	55.35405
Nestle Butterfinger	0.767	70.73564
Nestle Crunch	0.767	66.47068
Nik L Nip	0.976	22.44534
Now & Later	0.325	39.44680
Payday	0.767	46.29660
Peanut M&Ms	0.651	69.48379
Pixie Sticks	0.023	37.72234
Pop Rocks	0.837	41.26551
Red vines	0.116	37.34852
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
Ring pop	0.965	35.29076
Rolo	0.860	65.71629
Root Beer Barrels	0.069	29.70369
Runts	0.279	42.84914
Sixlets	0.081	34.72200
Skittles original	0.220	63.08514
Skittles wildberry	0.220	55.10370
Nestle Smarties	0.976	37.88719
Smarties candy	0.116	45.99583
Snickers	0.651	76.67378
Snickers Crisper	0.651	59.52925
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

```
Sugar Babies                       0.767   33.43755
Sugar Daddy                        0.325   32.23100
Super Bubble                       0.116   27.30386
Swedish Fish                       0.755   54.86111
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
Trolli Sour Bites                  0.255   47.17323
Twix                               0.906   81.64291
Twizzlers                          0.116   45.46628
Warheads                           0.116   39.01190
Welch's Fruit Snacks               0.313   44.37552
Werther's Original Caramel         0.267   41.90431
Whoppers                           0.848   49.52411
```

```
candy["Peanut butter M&M's", ]$winpercent
```

```
[1] 71.46505
```

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
```

The %in% operator is useful for checking the intersection of two vectors

```
c("barry", "liz", "chandra") %in% c("paul", "alice", "liz")
```

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

There is a useful skim() function in the skimr package that can help give you a quick overview of a given dataset. Let's install this package and try it on our candy data.

```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	

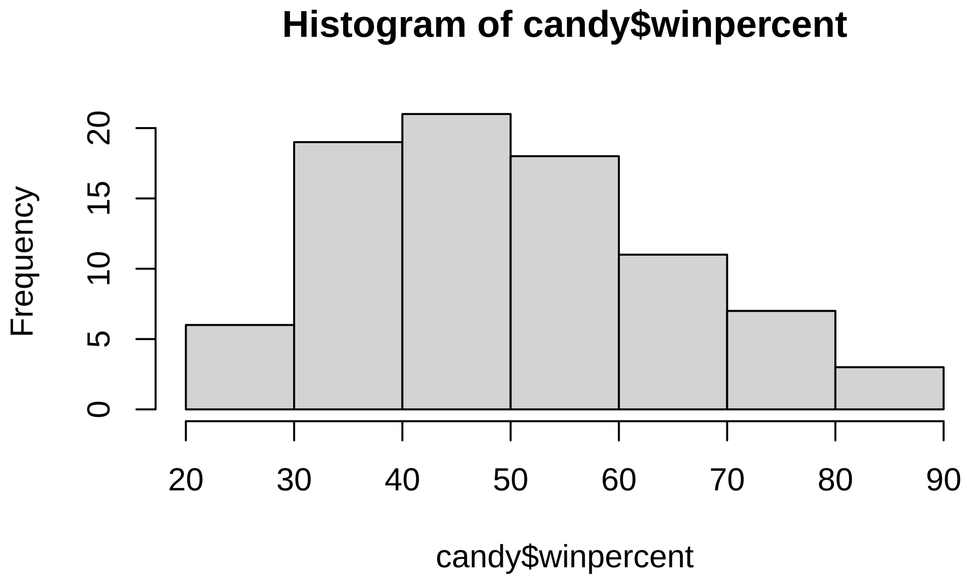
Q6. Is there any variable/column that looks to be on a different scale to the majority of the other columns in the dataset?

winpercent, sugarpercent, and pricepercent seem to be in percentage rather than a 0,1.

Q7. What do you think a zero and one represent for the candy$chocolate column?
The zero and one represent a true or false statement. 0=FALSE, 1=TRUE. Candies with a 1 in the chocolate column means that there is indeed chocolate.

Q8. Plot a histogram of winpercent values

```
hist(candy$winpercent)
```



Q9. Is the distribution of winpercent values symmetrical? No, it leans towards the left

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

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

```
mean(candy$chocolate)
```

```
[1] 0.4352941
```

```
mean(candy$fruity)
```

```
[1] 0.4470588
```

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

```r
choco.win<-as.logical(candy$chocolate)
candy[inds,]$winpercent
```

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

```r
fruit.win<-as.logical(candy$fruity)
candy[inds,]$winpercent
```

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

```
summary(fruit.win)
```

```
   Mode   FALSE    TRUE 
logical      47      38 
```

```
summary(choco.win)
```

```
   Mode   FALSE    TRUE 
logical      48      37 
```

Q12. Is this difference statistically significant?

not significant

```
t.test(fruit.win, choco.win)
```

```


	Welch Two Sample t-test

data:  fruit.win and choco.win
t = 0.15357, df = 168, p-value = 0.8781
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1394786  0.1630081
sample estimates:
mean of x mean of y 
0.4470588 0.4352941 
```

Overall Candy Rankings

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

There are two related functions that are useful here `sort()` and `order()`

So the 5 least are: Nik L Nip
Boston Baked Beans
Chiclets
Super Bubble
Jawbusters

```r
inds<-order(candy$winpercent)
head(candy[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 Peanut Butter cup
Reese’s Miniatures
Twix

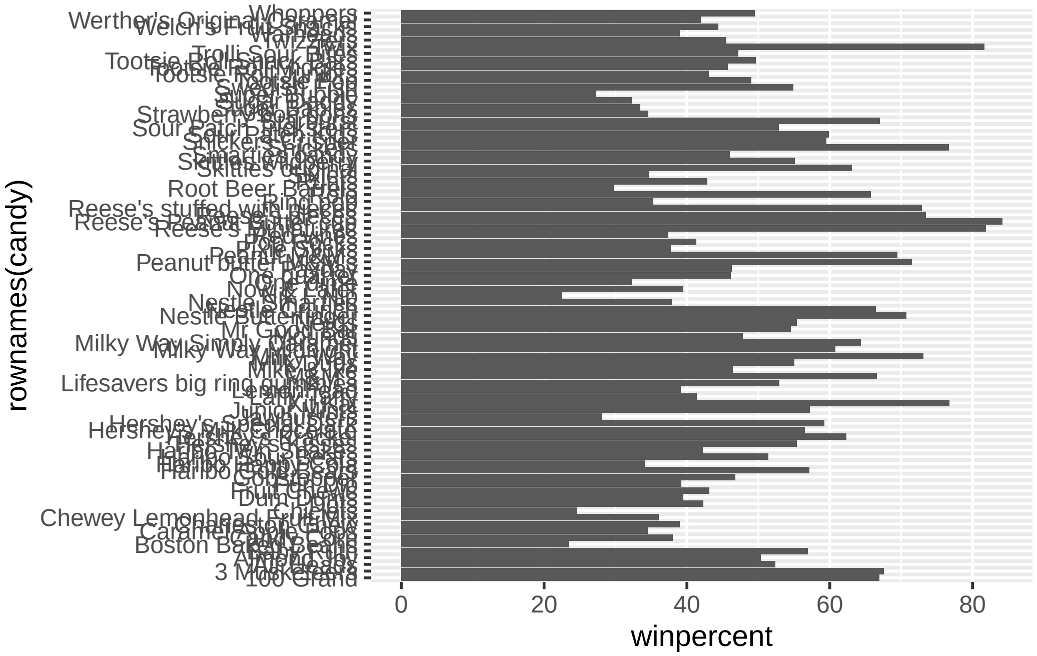
Kit Kat Snickers
Reese's pieces

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

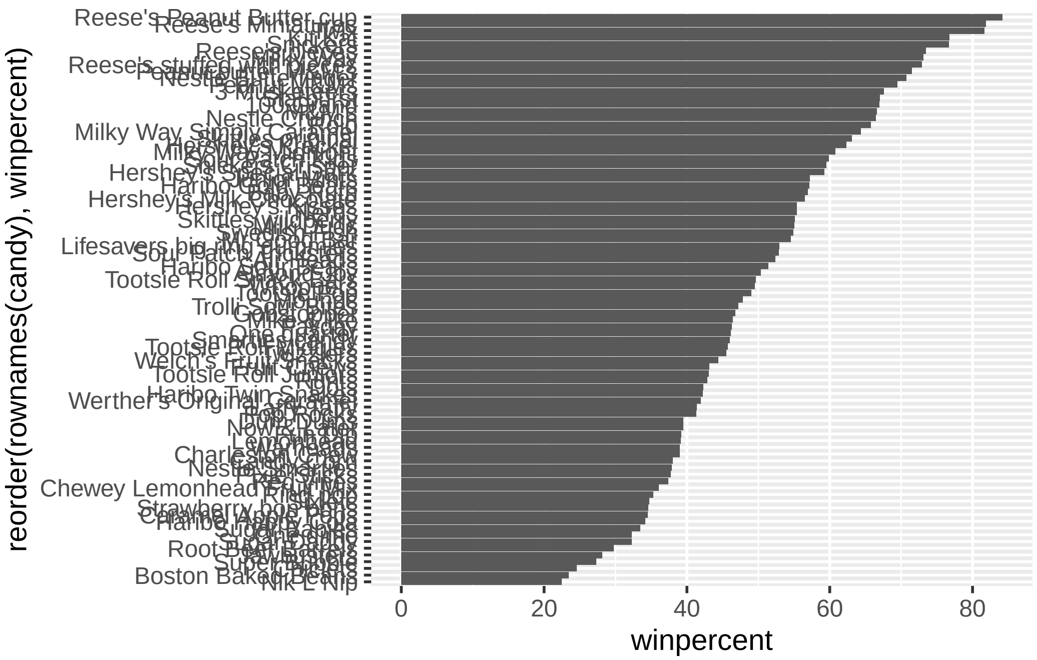
Q15. Make a first barplot of candy ranking based on winpercent values.

```
library(ggplot2)
ggplot(candy) +
  aes(winpercent, rownames(candy)) +
  geom_col()
```



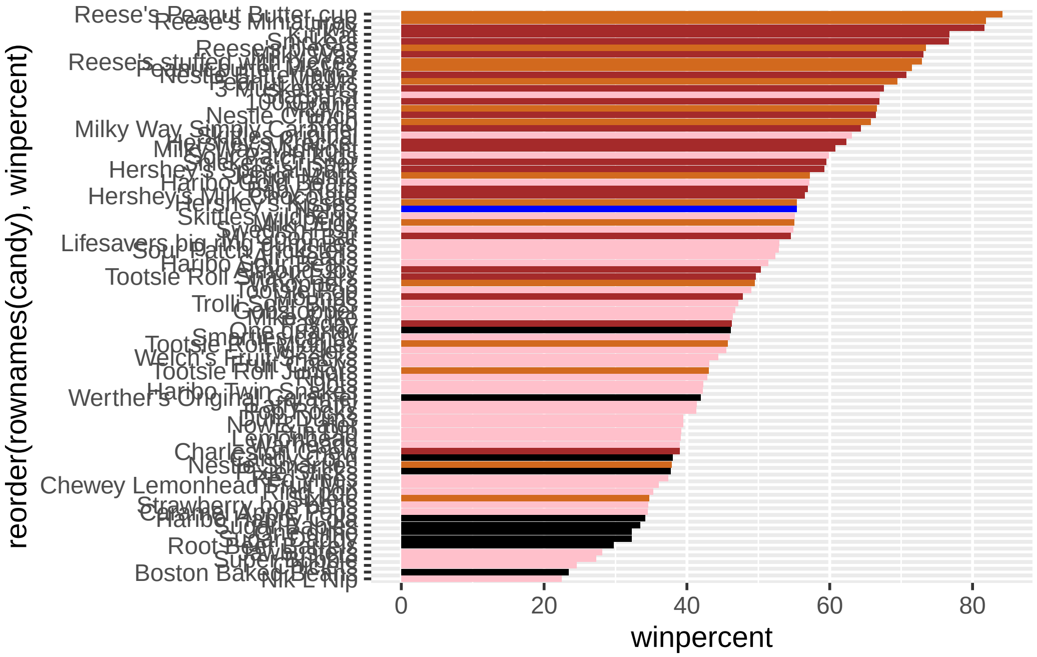
Q16. This is quite ugly, use the reorder() function to get the bars sorted by winpercent?

```
library(ggplot2)
ggplot(candy) +
  aes(winpercent, reorder(rownames(candy), winpercent)) +
  geom_col()
```



```r
my_cols=rep("black", nrow(candy))
my_cols[as.logical(candy$chocolate)] = "chocolate"
my_cols[as.logical(candy$bar)] = "brown"
my_cols[as.logical(candy$fruity)] = "pink"
my_cols[rownames(candy)=="Nerds"]<-"blue"
```

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



Q17. What is the worst ranked chocolate candy? Q18. What is the best ranked fruity candy?

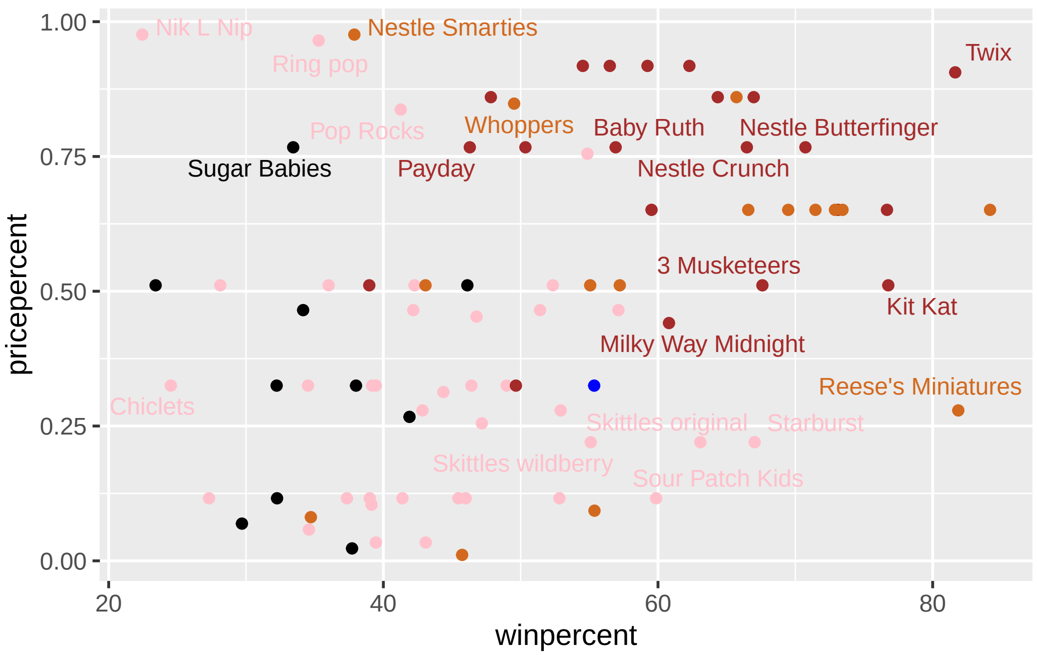
Worst: Nik L Nip; Best: Reeses Peanut Butter Cup

#Taking a look at pricepercent

What is the the best candy for the least money? One way to get at this would be to make a plot of winpercent vs the pricepercent variable. The pricepercent variable records the percentile rank of the candy's price against all the other candies in the dataset. Lower values are less expensive and high values more expensive.

```r
library(ggrepel)

# How about a plot of price vs win
ggplot(candy) +
  aes(winpercent, pricepercent, label=rownames(candy)) +
  geom_point(col=my_cols) +
  geom_text_repel(col=my_cols, size=3.3, max.overlaps = 5)
```

```
Warning: ggrepel: 65 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 Miniature

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

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

```
                     pricepercent winpercent
Tootsie Roll Midgies        0.011   45.73675
Pixie Sticks                0.023   37.72234
Dum Dums                    0.034   39.46056
Fruit Chews                 0.034   43.08892
Strawberry bon bons         0.058   34.57899
```

```
ordlow <- order(candy$pricepercent, decreasing = TRUE)
head( candy[ordlow,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
```

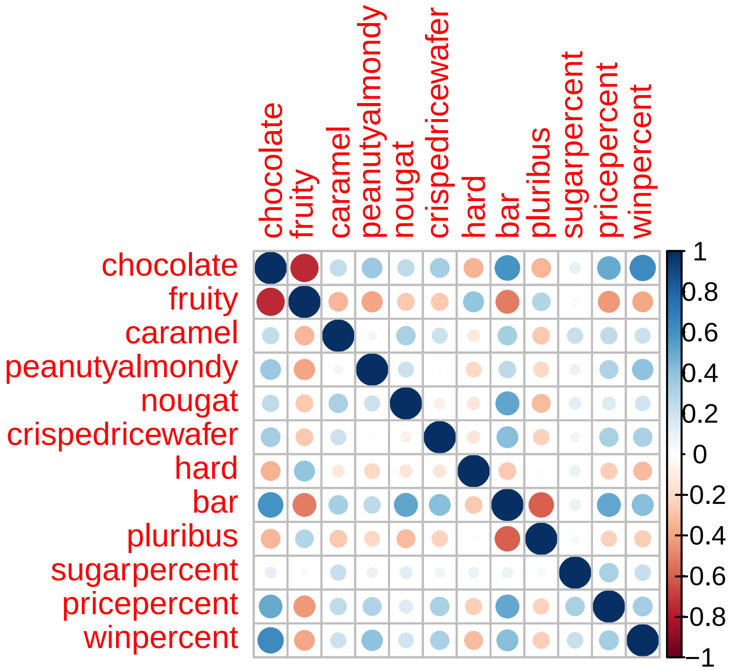
#Exploring the Correlation Structure

Now that we've explored the dataset a little, we'll see how the variables interact with one another. We'll use correlation and view the results with the corrplot package to plot a correlation matrix.

```r
library(corrplot)
```

```
corrplot 0.95 loaded
```

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



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

Chocolate fruity

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

Chocolate winpercent

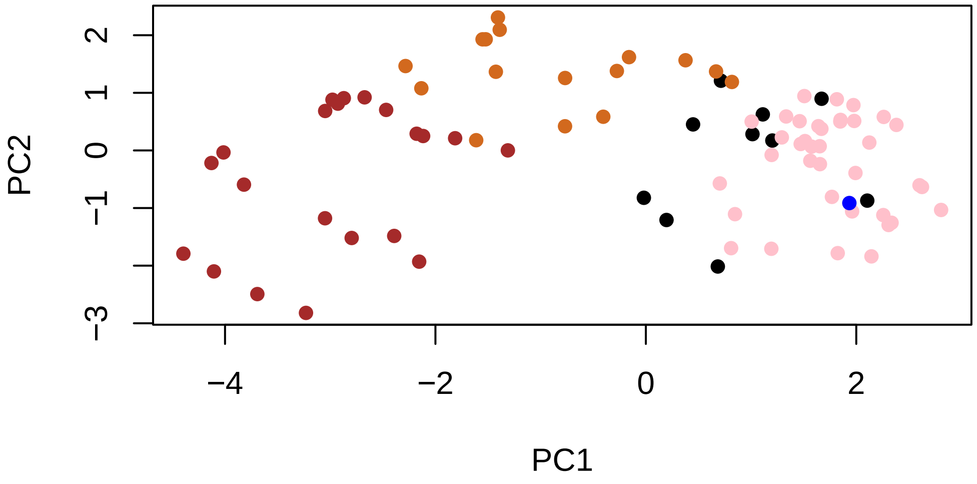
#Principal Component Analysis

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

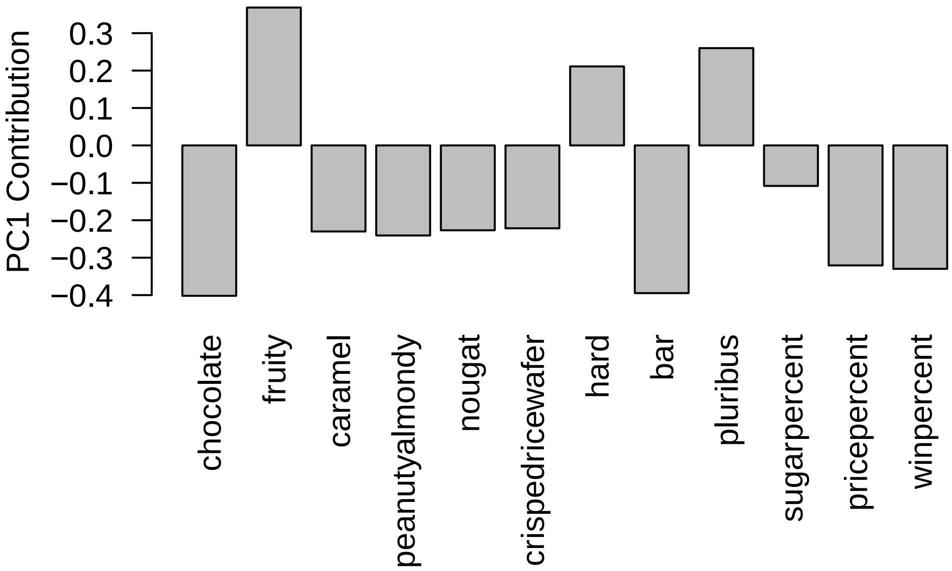
```
pca<-prcomp(candy, scale=TRUE)

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
```

```
plot(pca$x[,1:2], col=my_cols, pch=16)
```



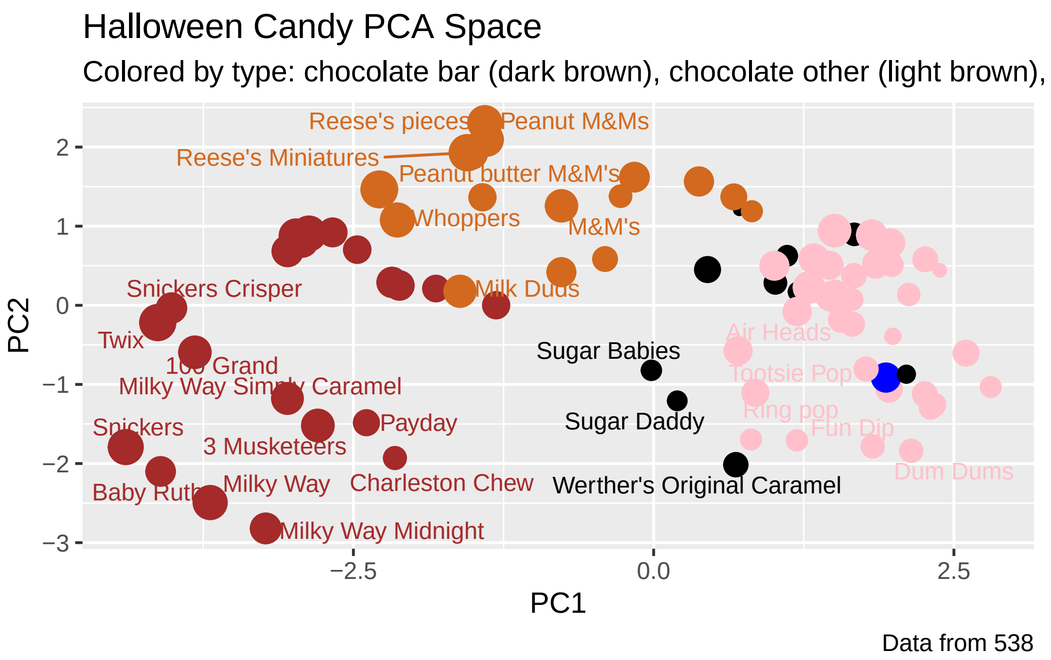
```r
par(mar=c(8,4,2,2))
barplot(pca$rotation[,1], las=2, ylab="PC1 Contribution")
```



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

```r
library(ggrepel)

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

p + geom_text_repel(size=3.3, col=my_cols, 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
```



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

Fruity, hard, pluribus. Yes because they are not correlated with the other factors. Its not in the same side as the other variables.