Class 10: Halloween mini project

Joshua Martin (PID: A18545389)

Table of contents

As it is nearly Halloween and the half way point in the quarter let’s do a mini project to help us figure out the best candy!

Our data comes from the 538 website and is available as a CSV file:

Data Import

```r
candy <- read.csv("candy-data.csv")
candy
```

```
                competitorname chocolate fruity caramel peanutyalmondy nougat
1                    100 Grand         1      0       1              0      0
2                 3 Musketeers         1      0       0              0      1
3                     One dime         0      0       0              0      0
4                  One quarter         0      0       0              0      0
5                    Air Heads         0      1       0              0      0
6                   Almond Joy         1      0       0              1      0
7                    Baby Ruth         1      0       1              1      1
8           Boston Baked Beans         0      0       0              1      0
9                   Candy Corn         0      0       0              0      0
10          Caramel Apple Pops         0      1       1              0      0
```

```
11             Charleston Chew         1      0       0              0      1
12  Chewey Lemonhead Fruit Mix         0      1       0              0      0
13                    Chiclets         0      1       0              0      0
14                        Dots         0      1       0              0      0
15                    Dum Dums         0      1       0              0      0
16                 Fruit Chews         0      1       0              0      0
17                     Fun Dip         0      1       0              0      0
18                  Gobstopper         0      1       0              0      0
19           Haribo Gold Bears         0      1       0              0      0
20           Haribo Happy Cola         0      0       0              0      0
21           Haribo Sour Bears         0      1       0              0      0
22          Haribo Twin Snakes         0      1       0              0      0
23            Hershey's Kisses         1      0       0              0      0
24           Hershey's Krackel         1      0       0              0      0
25    Hershey's Milk Chocolate         1      0       0              0      0
26      Hershey's Special Dark         1      0       0              0      0
27                  Jawbusters         0      1       0              0      0
28                Junior Mints         1      0       0              0      0
29                     Kit Kat         1      0       0              0      0
30                 Laffy Taffy         0      1       0              0      0
31                   Lemonhead         0      1       0              0      0
32 Lifesavers big ring gummies         0      1       0              0      0
33         Peanut butter M&M's         1      0       0              1      0
34                       M&M's         1      0       0              0      0
35                  Mike & Ike         0      1       0              0      0
36                   Milk Duds         1      0       1              0      0
37                   Milky Way         1      0       1              0      1
38          Milky Way Midnight         1      0       1              0      1
39    Milky Way Simply Caramel         1      0       1              0      0
40                      Mounds         1      0       0              0      0
41                 Mr Good Bar         1      0       0              1      0
42                       Nerds         0      1       0              0      0
43         Nestle Butterfinger         1      0       0              1      0
44               Nestle Crunch         1      0       0              0      0
45                   Nik L Nip         0      1       0              0      0
46                 Now & Later         0      1       0              0      0
47                      Payday         0      0       0              1      1
48                 Peanut M&Ms         1      0       0              1      0
49                Pixie Sticks         0      0       0              0      0
50                   Pop Rocks         0      1       0              0      0
51                   Red vines         0      1       0              0      0
52          Reese's Miniatures         1      0       0              1      0
53   Reese's Peanut Butter cup         1      0       0              1      0
```

```
54              Reese's pieces         1      0       0              1      0
55 Reese's stuffed with pieces         1      0       0              1      0
56                    Ring pop         0      1       0              0      0
57                        Rolo         1      0       1              0      0
58           Root Beer Barrels         0      0       0              0      0
59                       Runts         0      1       0              0      0
60                     Sixlets         1      0       0              0      0
61           Skittles original         0      1       0              0      0
62          Skittles wildberry         0      1       0              0      0
63             Nestle Smarties         1      0       0              0      0
64              Smarties candy         0      1       0              0      0
65                    Snickers         1      0       1              1      1
66            Snickers Crisper         1      0       1              1      0
67             Sour Patch Kids         0      1       0              0      0
68       Sour Patch Tricksters         0      1       0              0      0
69                   Starburst         0      1       0              0      0
70         Strawberry bon bons         0      1       0              0      0
71                Sugar Babies         0      0       1              0      0
72                 Sugar Daddy         0      0       1              0      0
73                Super Bubble         0      1       0              0      0
74                Swedish Fish         0      1       0              0      0
75                 Tootsie Pop         1      1       0              0      0
76        Tootsie Roll Juniors         1      0       0              0      0
77        Tootsie Roll Midgies         1      0       0              0      0
78     Tootsie Roll Snack Bars         1      0       0              0      0
79           Trolli Sour Bites         0      1       0              0      0
80                        Twix         1      0       1              0      0
81                   Twizzlers         0      1       0              0      0
82                    Warheads         0      1       0              0      0
83        Welch's Fruit Snacks         0      1       0              0      0
84  Werther's Original Caramel         0      0       1              0      0
85                    Whoppers         1      0       0              0      0
   crispedricewafer hard bar pluribus sugarpercent pricepercent winpercent
1                 1    0   1        0        0.732        0.860   66.97173
2                 0    0   1        0        0.604        0.511   67.60294
3                 0    0   0        0        0.011        0.116   32.26109
4                 0    0   0        0        0.011        0.511   46.11650
5                 0    0   0        0        0.906        0.511   52.34146
6                 0    0   1        0        0.465        0.767   50.34755
7                 0    0   1        0        0.604        0.767   56.91455
8                 0    0   0        1        0.313        0.511   23.41782
9                 0    0   0        1        0.906        0.325   38.01096
10                0    0   0        0        0.604        0.325   34.51768
```

11	0	0	1	0	0.604	0.511	38.97504
12	0	0	0	1	0.732	0.511	36.01763
13	0	0	0	1	0.046	0.325	24.52499
14	0	0	0	1	0.732	0.511	42.27208
15	0	1	0	0	0.732	0.034	39.46056
16	0	0	0	1	0.127	0.034	43.08892
17	0	1	0	0	0.732	0.325	39.18550
18	0	1	0	1	0.906	0.453	46.78335
19	0	0	0	1	0.465	0.465	57.11974
20	0	0	0	1	0.465	0.465	34.15896
21	0	0	0	1	0.465	0.465	51.41243
22	0	0	0	1	0.465	0.465	42.17877
23	0	0	0	1	0.127	0.093	55.37545
24	1	0	1	0	0.430	0.918	62.28448
25	0	0	1	0	0.430	0.918	56.49050
26	0	0	1	0	0.430	0.918	59.23612
27	0	1	0	1	0.093	0.511	28.12744
28	0	0	0	1	0.197	0.511	57.21925
29	1	0	1	0	0.313	0.511	76.76860
30	0	0	0	0	0.220	0.116	41.38956
31	0	1	0	0	0.046	0.104	39.14106
32	0	0	0	0	0.267	0.279	52.91139
33	0	0	0	1	0.825	0.651	71.46505
34	0	0	0	1	0.825	0.651	66.57458
35	0	0	0	1	0.872	0.325	46.41172
36	0	0	0	1	0.302	0.511	55.06407
37	0	0	1	0	0.604	0.651	73.09956
38	0	0	1	0	0.732	0.441	60.80070
39	0	0	1	0	0.965	0.860	64.35334
40	0	0	1	0	0.313	0.860	47.82975
41	0	0	1	0	0.313	0.918	54.52645
42	0	1	0	1	0.848	0.325	55.35405
43	0	0	1	0	0.604	0.767	70.73564
44	1	0	1	0	0.313	0.767	66.47068
45	0	0	0	1	0.197	0.976	22.44534
46	0	0	0	1	0.220	0.325	39.44680
47	0	0	1	0	0.465	0.767	46.29660
48	0	0	0	1	0.593	0.651	69.48379
49	0	0	0	1	0.093	0.023	37.72234
50	0	1	0	1	0.604	0.837	41.26551
51	0	0	0	1	0.581	0.116	37.34852
52	0	0	0	0	0.034	0.279	81.86626
53	0	0	0	0	0.720	0.651	84.18029

```
54                0    0   0         1       0.406        0.651   73.43499
55                0    0   0         0       0.988        0.651   72.88790
56                0    1   0         0       0.732        0.965   35.29076
57                0    0   0         1       0.860        0.860   65.71629
58                0    1   0         1       0.732        0.069   29.70369
59                0    1   0         1       0.872        0.279   42.84914
60                0    0   0         1       0.220        0.081   34.72200
61                0    0   0         1       0.941        0.220   63.08514
62                0    0   0         1       0.941        0.220   55.10370
63                0    0   0         1       0.267        0.976   37.88719
64                0    1   0         1       0.267        0.116   45.99583
65                0    0   1         0       0.546        0.651   76.67378
66                1    0   1         0       0.604        0.651   59.52925
67                0    0   0         1       0.069        0.116   59.86400
68                0    0   0         1       0.069        0.116   52.82595
69                0    0   0         1       0.151        0.220   67.03763
70                0    1   0         1       0.569        0.058   34.57899
71                0    0   0         1       0.965        0.767   33.43755
72                0    0   0         0       0.418        0.325   32.23100
73                0    0   0         0       0.162        0.116   27.30386
74                0    0   0         1       0.604        0.755   54.86111
75                0    1   0         0       0.604        0.325   48.98265
76                0    0   0         0       0.313        0.511   43.06890
77                0    0   0         1       0.174        0.011   45.73675
78                0    0   1         0       0.465        0.325   49.65350
79                0    0   0         1       0.313        0.255   47.17323
80                1    0   1         0       0.546        0.906   81.64291
81                0    0   0         0       0.220        0.116   45.46628
82                0    1   0         0       0.093        0.116   39.01190
83                0    0   0         1       0.313        0.313   44.37552
84                0    1   0         0       0.186        0.267   41.90431
85                1    0   0         1       0.872        0.848   49.52411
```

```r
library(flextable)
flextable::flextable(head(candy, 10))
```

competitorname	chocolate	fruity	caramel	peanutyalmondy	nougat	crispedricewafer	hard	bar
100 Grand	1	0	1	0	0	1	0	1
3 Musketeers	1	0	0	0	1	0	0	1

competitorname	chocolate	fruity	caramel	peanutyalmondy	nougat	crispedricewafer	hard	bar
One dime	0	0	0	0	0	0	0	0
One quarter	0	0	0	0	0	0	0	0
Air Heads	0	1	0	0	0	0	0	0
Almond Joy	1	0	0	1	0	0	0	1
Baby Ruth	1	0	1	1	1	0	0	1
Boston Baked Beans	0	0	0	1	0	0	0	0
Candy Corn	0	0	0	0	0	0	0	0
Caramel Apple Pops	0	1	1	0	0	0	0	0

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

```
nrow(candy)
```

```
[1] 85
```

```
candy |>
  nrow()
```

```
[1] 85
```

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.1     v stringr   1.5.2
v ggplot2   4.0.0     v tibble    3.3.0
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.1.0
```

```
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x purrr::compose() masks flextable::compose()
x dplyr::filter()  masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```
candy %>%
  nrow()
```

```
[1] 85
```

```
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
```

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

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

```
[1] 38
```

```
candy |>
  nrow()
```

```
[1] 85
```

```r
library(tidyverse)
candy %>%
  nrow()
```

```
[1] 85
```

```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
```

2. What is your favorite candy?

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

My favorite candy in the dataset is Skittles

```r
candy["Skittles original", ]$winpercent
```

```
[1] 63.08514
```

```r
library(dplyr)
candy |>
  filter(rownames(candy)=="Twix") |>
  select(winpercent)
```

```
     winpercent
Twix   81.64291
```

```
candy |>
  filter(rownames(candy)=="Almond Joy") |>
  select(winpercent)
```

```
           winpercent
Almond Joy   50.34755
```

Quick overview of the dataset

```
skimr::skim(candy)
```

Table 2: 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	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
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	

Q4. What is the winpercent value for “Kit Kat”?

```
candy["Twix", ]$winpercent
```

```
[1] 81.64291
```

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

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

```
[1] 49.6535
```

```
library("skimr")
skim(candy)
```

Table 4: 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	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
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?

```
skim(candy)
```

Table 6: 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	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist

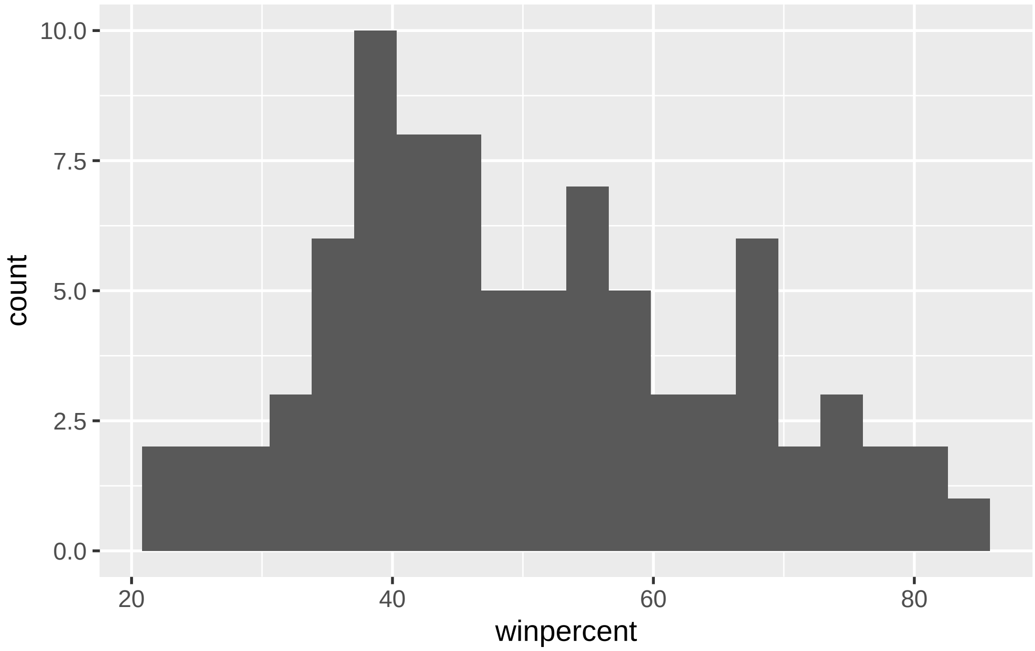
The variable winpercent stands out because the rest of mostly binary 0s and 1s. Winpercent represents a popularity score and is not categorical and thus is on a different numerical scale than the rest. The winpercent is on 0-100 scale the rest are 0-1 scale.

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

That the data does not contain chocolate.

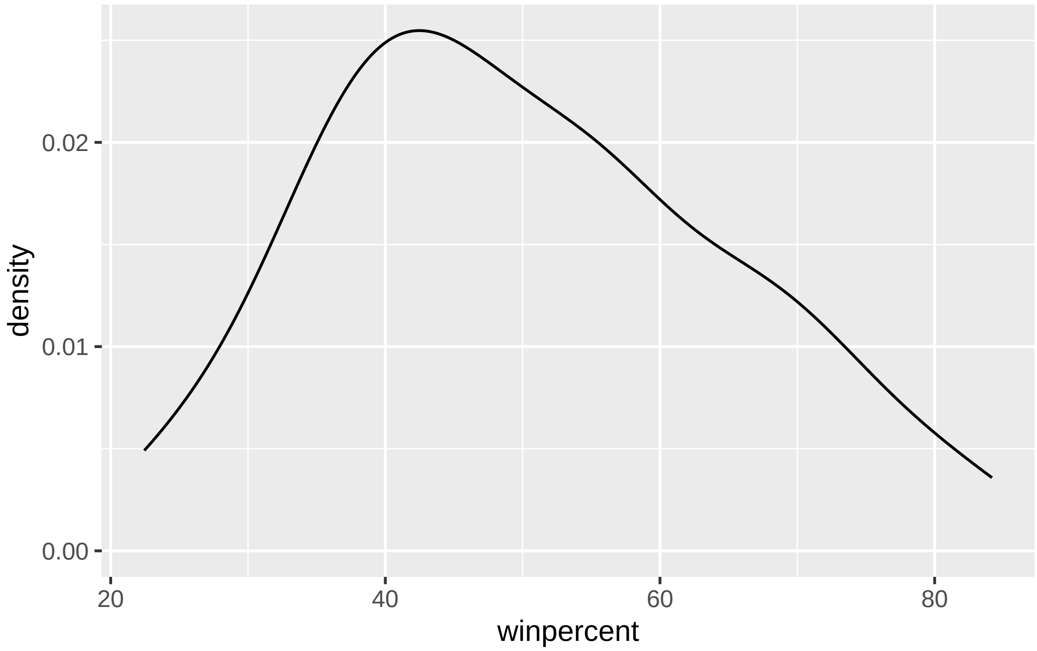
> Q8. Plot a histogram of winpercent values

```r
library(ggplot2)

ggplot(candy) +
  aes(winpercent) +
  geom_histogram(bins=20)
```



> Q9. Is the distribution of winpercent values symmetrical?

```r
ggplot(candy) +
  aes(winpercent) +
  geom_density()
```



```r
mean(candy$winpercent)
```

```
[1] 50.31676
```

```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?

```r
# 1. Find all chocolate candy in the dataset
# 2. Find their winpercent values
# 3. Calculate the mean of these values

# 4-6. Do the same for fruit candy
# 7. Compare mean winpercents of chocolate vs fruity
# 8. Pick the highest as the winner

choc.inds <- candy$chocolate==1
choc.win <- candy[choc.inds, ]$winpercent
choc.mean <- mean(choc.win)
choc.mean
```

```
[1] 60.92153
```

```
mean(candy[candy$chocolate==1,]$winpercent)
```

```
[1] 60.92153
```

```
fruity.inds <- candy$fruit==1
fruity.win <- candy[fruity.inds, ]$winpercent
fruity.mean <- mean(fruity.win)
fruity.mean
```

```
[1] 44.11974
```

```
mean(candy[candy$fruity==1,]$winpercent)
```

```
[1] 44.11974
```

```
candy |>
  filter(chocolate==1) |>
  select(winpercent)
```

```
                            winpercent
100 Grand                     66.97173
3 Musketeers                  67.60294
Almond Joy                    50.34755
Baby Ruth                     56.91455
Charleston Chew               38.97504
Hershey's Kisses              55.37545
Hershey's Krackel             62.28448
Hershey's Milk Chocolate      56.49050
```

```
Hershey's Special Dark        59.23612
Junior Mints                  57.21925
Kit Kat                       76.76860
Peanut butter M&M's           71.46505
M&M's                         66.57458
Milk Duds                     55.06407
Milky Way                     73.09956
Milky Way Midnight            60.80070
Milky Way Simply Caramel      64.35334
Mounds                        47.82975
Mr Good Bar                   54.52645
Nestle Butterfinger           70.73564
Nestle Crunch                 66.47068
Peanut M&Ms                   69.48379
Reese's Miniatures            81.86626
Reese's Peanut Butter cup     84.18029
Reese's pieces                73.43499
Reese's stuffed with pieces   72.88790
Rolo                          65.71629
Sixlets                       34.72200
Nestle Smarties               37.88719
Snickers                      76.67378
Snickers Crisper              59.52925
Tootsie Pop                   48.98265
Tootsie Roll Juniors          43.06890
Tootsie Roll Midgies          45.73675
Tootsie Roll Snack Bars       49.65350
Twix                          81.64291
Whoppers                      49.52411
```

```
candy |>
  filter(fruity==1) |>
  select(winpercent)
```

```
                            winpercent
Air Heads                     52.34146
Caramel Apple Pops            34.51768
Chewey Lemonhead Fruit Mix    36.01763
Chiclets                      24.52499
Dots                          42.27208
Dum Dums                      39.46056
Fruit Chews                   43.08892
```

```
Fun Dip                      39.18550
Gobstopper                   46.78335
Haribo Gold Bears            57.11974
Haribo Sour Bears            51.41243
Haribo Twin Snakes           42.17877
Jawbusters                   28.12744
Laffy Taffy                  41.38956
Lemonhead                    39.14106
Lifesavers big ring gummies  52.91139
Mike & Ike                   46.41172
Nerds                        55.35405
Nik L Nip                    22.44534
Now & Later                  39.44680
Pop Rocks                    41.26551
Red vines                    37.34852
Ring pop                     35.29076
Runts                        42.84914
Skittles original            63.08514
Skittles wildberry           55.10370
Smarties candy               45.99583
Sour Patch Kids              59.86400
Sour Patch Tricksters        52.82595
Starburst                    67.03763
Strawberry bon bons          34.57899
Super Bubble                 27.30386
Swedish Fish                 54.86111
Tootsie Pop                  48.98265
Trolli Sour Bites            47.17323
Twizzlers                    45.46628
Warheads                     39.01190
Welch's Fruit Snacks         44.37552
```

Q12. Is this differece statistically significant?

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

```

    Welch Two Sample t-test

data:  choc.win and fruity.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
```

3. Overall Candy Ranking

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

Nik L Nip Boston Baked Beans Chiclets Super Bubble Jawbusters

```r
head(candy[order(candy$winpercent),], n=5)
```

```
                   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
                   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
                   winpercent
Nik L Nip            22.44534
Boston Baked Beans   23.41782
Chiclets             24.52499
Super Bubble         27.30386
Jawbusters           28.12744
```

```r
candy %>% arrange(winpercent) %>% head(5)
```

```
                   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
```

```
                   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
                   winpercent
Nik L Nip            22.44534
Boston Baked Beans   23.41782
Chiclets             24.52499
Super Bubble         27.30386
Jawbusters           28.12744
```

```r
candy |>
  arrange(winpercent) |>
  head(5)
```

```
                   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
                   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
                   winpercent
Nik L Nip            22.44534
Boston Baked Beans   23.41782
Chiclets             24.52499
Super Bubble         27.30386
Jawbusters           28.12744
```

```r
x <- c(5,1,10,4)
#sort(x)
order(x)
```

```
[1] 2 4 1 3
```

```
#(candy$winpercent)
```

```
ord.ind <- order(candy$winpercent)
head(candy[ord.ind,],5)
```

```
                   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
                   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
                   winpercent
Nik L Nip            22.44534
Boston Baked Beans   23.41782
Chiclets             24.52499
Super Bubble         27.30386
Jawbusters           28.12744
```

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

```
candy |>
  arrange(winpercent) |>
  tail(5)
```

```
                          chocolate fruity caramel peanutyalmondy nougat
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
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
```

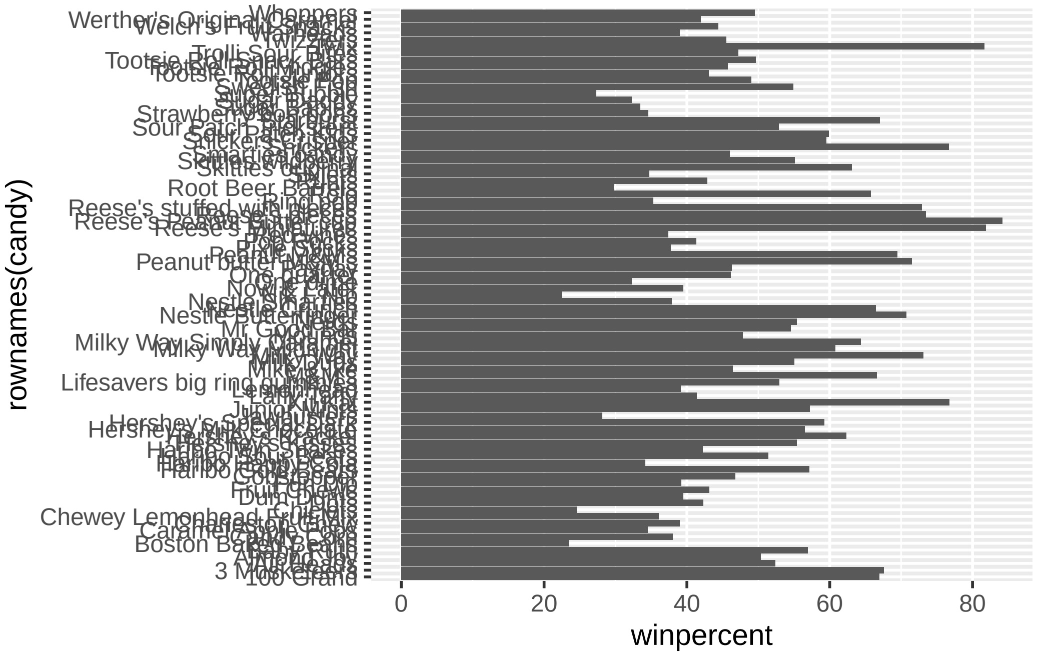
```
Reese's Peanut Butter cup                0    0   0        0        0.720
                          pricepercent winpercent
Snickers                         0.651   76.67378
Kit Kat                          0.511   76.76860
Twix                             0.906   81.64291
Reese's Miniatures               0.279   81.86626
Reese's Peanut Butter cup        0.651   84.18029
```

```
candy |>
  arrange(-winpercent) |>
  head(5)
```

```
                          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
                          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
                          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
```

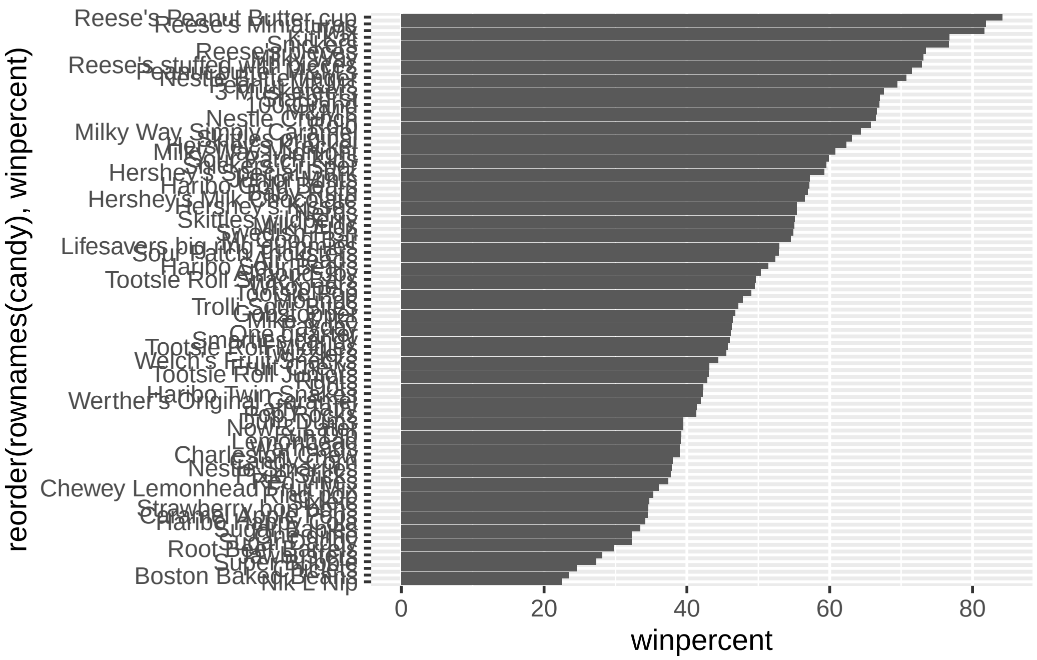
Q15. Make a first barplot of candy ranking based on winpercent values.

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



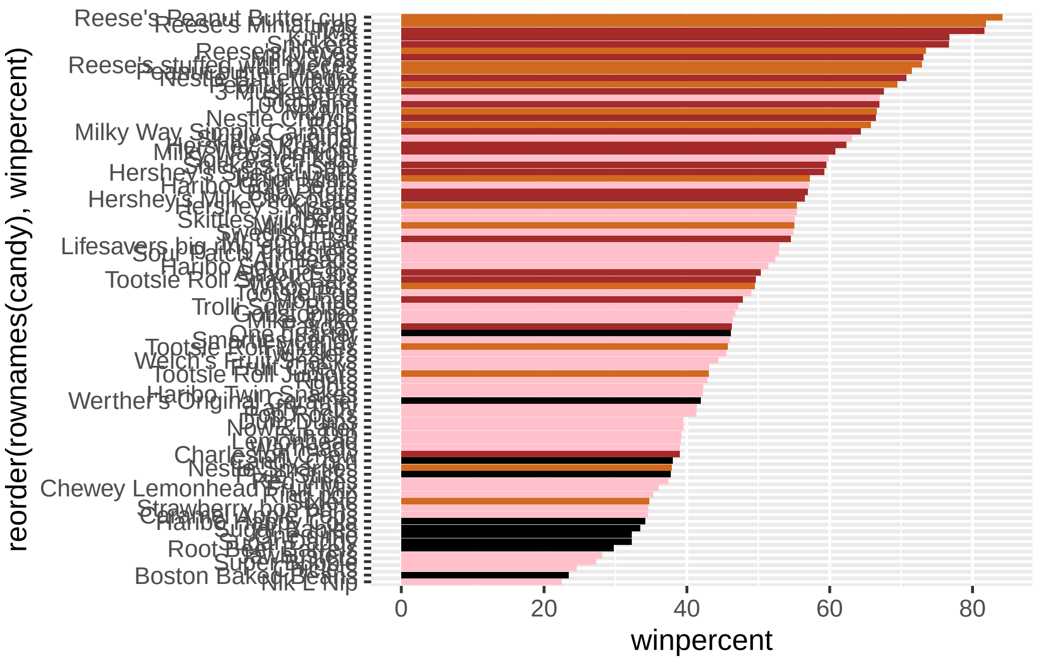
Q16. This is quite ugly, use the reorder() function to get the bars sorted by winpercent?

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



```
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"
```

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



Q17. What is the worst ranked chocolate candy?

Sixlets

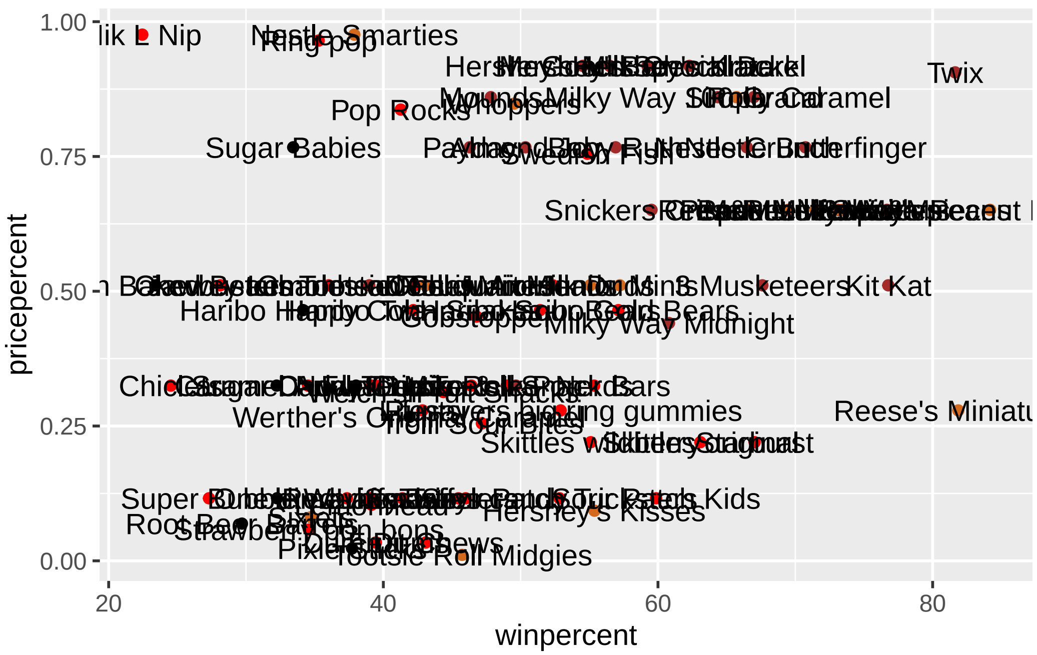
Q18. What is the best ranked fruity candy?

Nik L Nip

4. Winpercent and Pricepercent

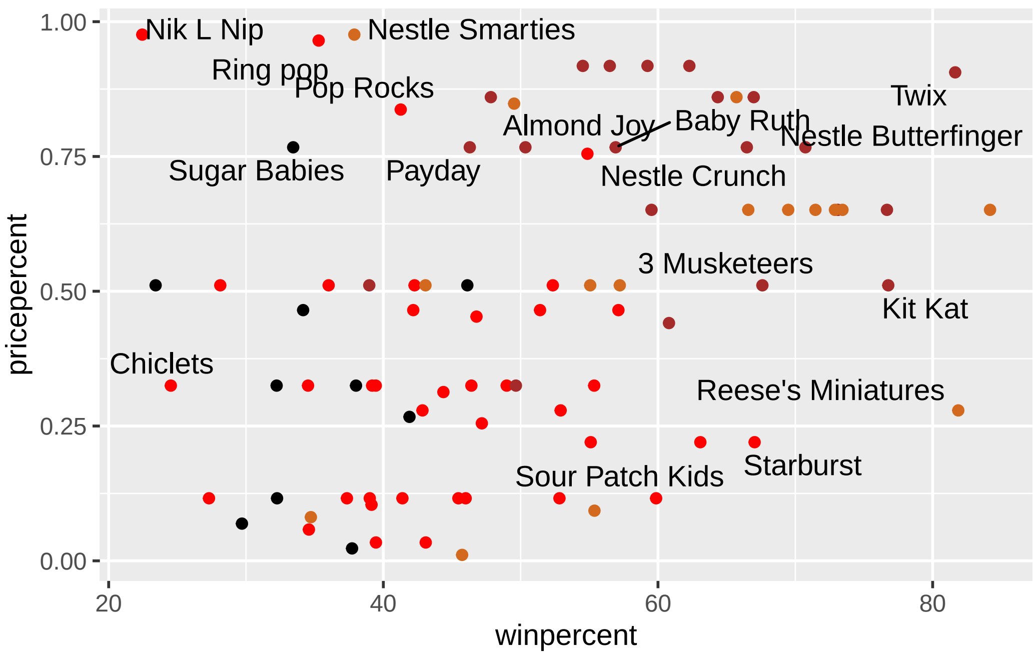
A plot with both variables/columns winpercent and pricepercent

```
my_cols[as.logical(candy$fruit)] <- "red"

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



```r
library(ggrepel)

ggplot(candy) +
  aes(x=winpercent,
      y=pricepercent,
      label=rownames(candy)) +
  geom_point(col=my_cols) +
  geom_text_repel(max.overlaps = 7)
```

```
Warning: ggrepel: 68 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```



Now that we've explored the dataset a little, we'll see how the variables interact with one another. We will plot a correlation matrix.

```
library(corrplot)
```

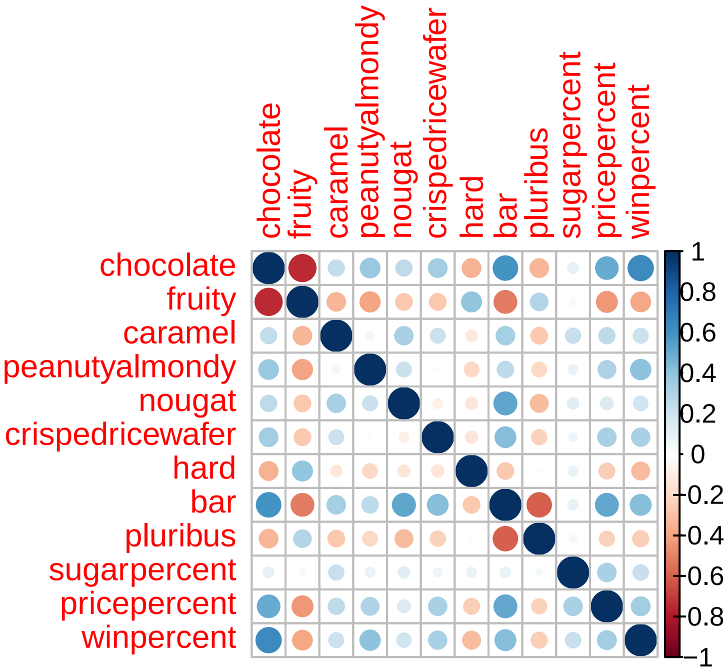
```
corrplot 0.95 loaded
```

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

```
                  chocolate      fruity     caramel peanutyalmondy      nougat
chocolate         1.0000000 -0.74172106  0.24987535     0.37782357  0.25489183
fruity           -0.7417211  1.00000000 -0.33548538    -0.39928014 -0.26936712
caramel           0.2498753 -0.33548538  1.00000000     0.05935614  0.32849280
peanutyalmondy    0.3778236 -0.39928014  0.05935614     1.00000000  0.21311310
nougat            0.2548918 -0.26936712  0.32849280     0.21311310  1.00000000
crispedricewafer  0.3412098 -0.26936712  0.21311310    -0.01764631 -0.08974359
hard             -0.3441769  0.39067750 -0.12235513    -0.20555661 -0.13867505
bar               0.5974211 -0.51506558  0.33396002     0.26041960  0.52297636
pluribus         -0.3396752  0.29972522 -0.26958501    -0.20610932 -0.31033884
sugarpercent      0.1041691 -0.03439296  0.22193335     0.08788927  0.12308135
pricepercent      0.5046754 -0.43096853  0.25432709     0.30915323  0.15319643
```

```
winpercent        0.6365167 -0.38093814  0.21341630     0.40619220  0.19937530
                 crispedricewafer        hard         bar    pluribus
chocolate              0.34120978 -0.34417691  0.59742114 -0.33967519
fruity                -0.26936712  0.39067750 -0.51506558  0.29972522
caramel                0.21311310 -0.12235513  0.33396002 -0.26958501
peanutyalmondy        -0.01764631 -0.20555661  0.26041960 -0.20610932
nougat                -0.08974359 -0.13867505  0.52297636 -0.31033884
crispedricewafer       1.00000000 -0.13867505  0.42375093 -0.22469338
hard                  -0.13867505  1.00000000 -0.26516504  0.01453172
bar                    0.42375093 -0.26516504  1.00000000 -0.59340892
pluribus              -0.22469338  0.01453172 -0.59340892  1.00000000
sugarpercent           0.06994969  0.09180975  0.09998516  0.04552282
pricepercent           0.32826539 -0.24436534  0.51840654 -0.22079363
winpercent             0.32467965 -0.31038158  0.42992933 -0.24744787
                 sugarpercent pricepercent winpercent
chocolate          0.10416906    0.5046754  0.6365167
fruity            -0.03439296   -0.4309685 -0.3809381
caramel            0.22193335    0.2543271  0.2134163
peanutyalmondy     0.08788927    0.3091532  0.4061922
nougat             0.12308135    0.1531964  0.1993753
crispedricewafer   0.06994969    0.3282654  0.3246797
hard               0.09180975   -0.2443653 -0.3103816
bar                0.09998516    0.5184065  0.4299293
pluribus           0.04552282   -0.2207936 -0.2474479
sugarpercent       1.00000000    0.3297064  0.2291507
pricepercent       0.32970639    1.0000000  0.3453254
winpercent         0.22915066    0.3453254  1.0000000
```

```
corrplot(cij)
```



5. Principal Component Analysis

The function to use i called `prcomp()` with an optional `scale=T/F` argument.

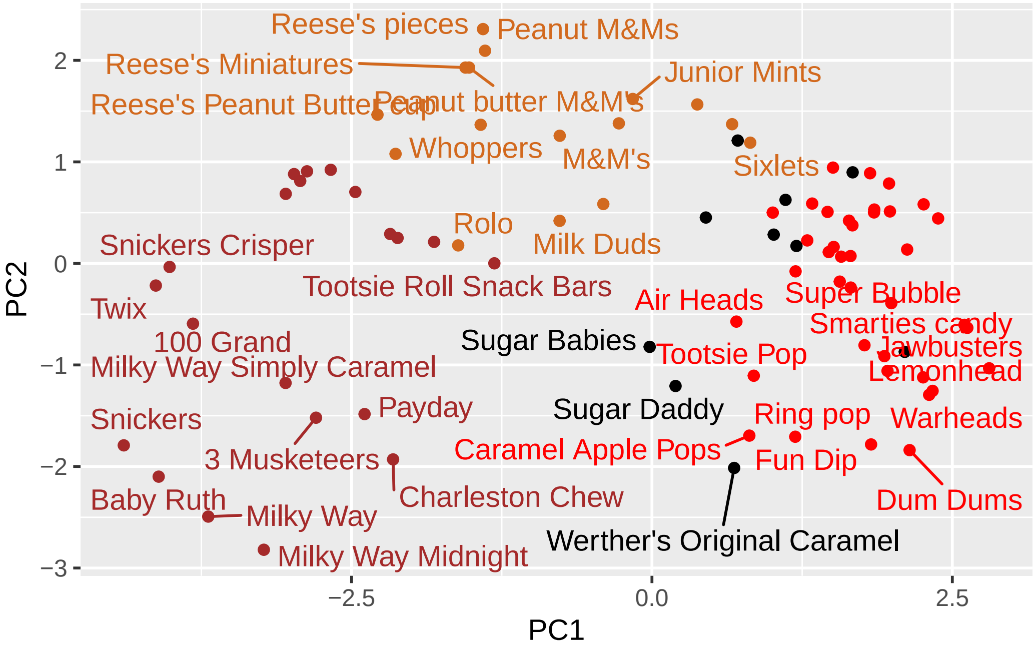
```
pca <- prcomp(candy, scale=TRUE)
summary(pca)
```

```
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
```

Our main PCA result figure

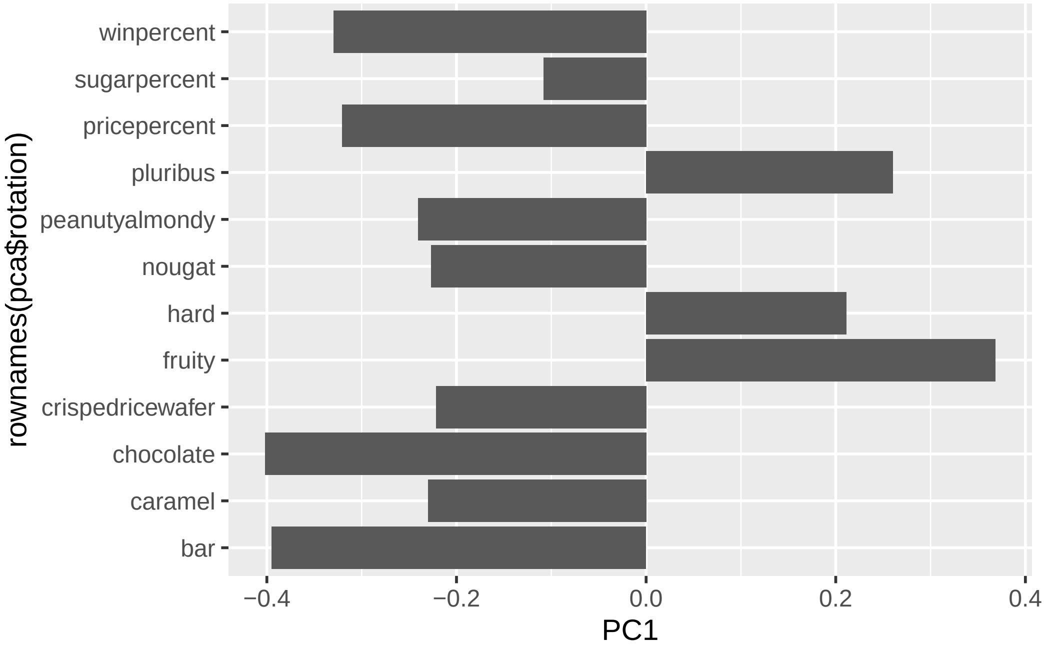
```
ggplot(pca$x) +
  aes(PC1, PC2, label=rownames(pca$x)) +
  geom_point(col=my_cols) +
  geom_text_repel(col=my_cols)
```

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



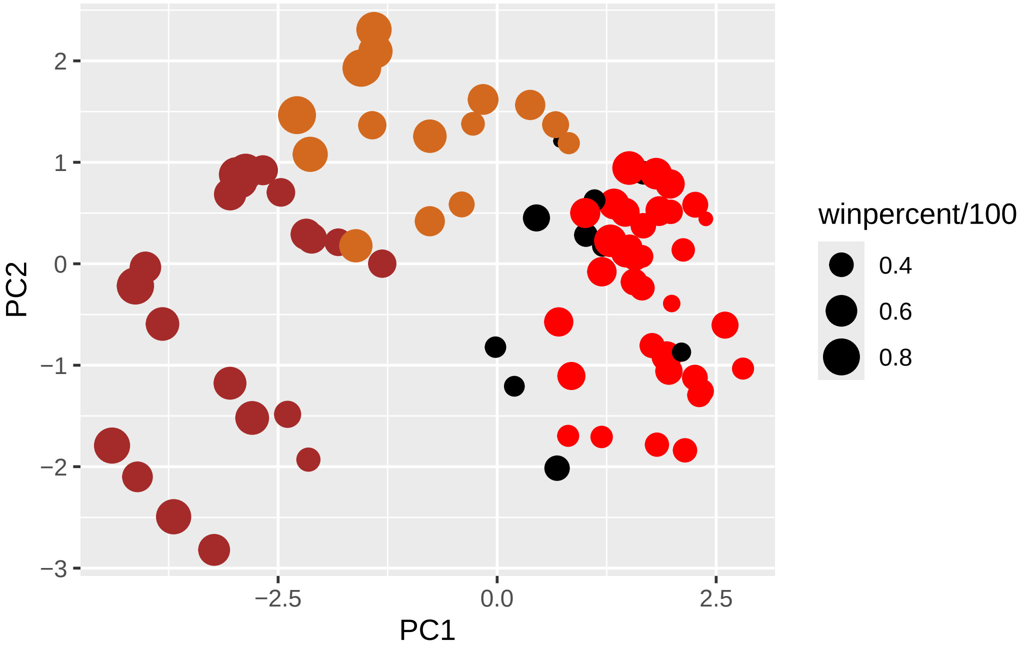
We should also examine the variable "loadings" or contrubutions of the orginional variables to the new PCs

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



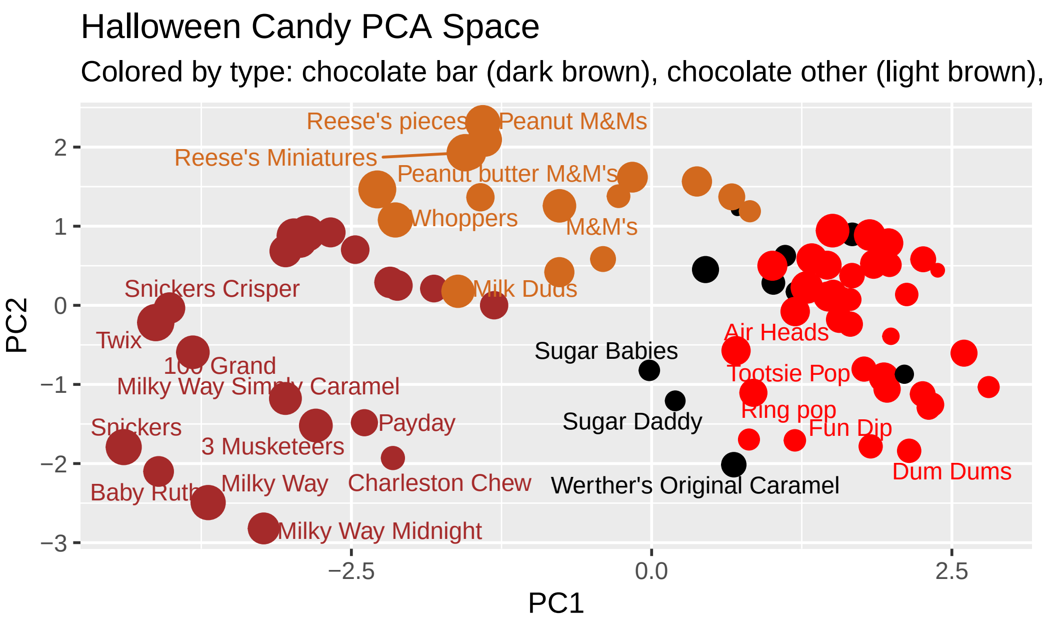
```r
my_data <- cbind(candy, pca$x[,1:3])
```

```r
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
```



```
library(ggrepel)

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
```



Data from 538

Interactive plots that can be zoomes on and "brushed" over can be made with the **plotly** package. It's output is interactive and will not render to PDF :-(

```
library(plotly)
```

```

Attaching package: 'plotly'

The following object is masked from 'package:ggplot2':

    last_plot

The following objects are masked from 'package:flextable':

    highlight, style

The following object is masked from 'package:stats':

    filter

The following object is masked from 'package:graphics':

    layout
```

```
#plotly(p)
```