Lab 10

Gretel Warmuth (PID: A17595945)

Importing candy data

Uploading candy data from 538 into this R project:

```
candy <- read.csv("candy-data.csv", row.names = 1)
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
```

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

```
nrow(candy)
```

```
[1] 85
```

```
rownames(candy)
```

```
 [1] "100 Grand"                 "3 Musketeers"
 [3] "One dime"                  "One quarter"
 [5] "Air Heads"                 "Almond Joy"
 [7] "Baby Ruth"                 "Boston Baked Beans"
 [9] "Candy Corn"                "Caramel Apple Pops"
[11] "Charleston Chew"           "Chewey Lemonhead Fruit Mix"
[13] "Chiclets"                  "Dots"
[15] "Dum Dums"                  "Fruit Chews"
[17] "Fun Dip"                   "Gobstopper"
[19] "Haribo Gold Bears"         "Haribo Happy Cola"
[21] "Haribo Sour Bears"         "Haribo Twin Snakes"
[23] "Hershey's Kisses"          "Hershey's Krackel"
[25] "Hershey's Milk Chocolate"  "Hershey's Special Dark"
```

```
[27] "Jawbusters"                   "Junior Mints"
[29] "Kit Kat"                      "Laffy Taffy"
[31] "Lemonhead"                    "Lifesavers big ring gummies"
[33] "Peanut butter M&M's"          "M&M's"
[35] "Mike & Ike"                   "Milk Duds"
[37] "Milky Way"                    "Milky Way Midnight"
[39] "Milky Way Simply Caramel"     "Mounds"
[41] "Mr Good Bar"                  "Nerds"
[43] "Nestle Butterfinger"          "Nestle Crunch"
[45] "Nik L Nip"                    "Now & Later"
[47] "Payday"                       "Peanut M&Ms"
[49] "Pixie Sticks"                 "Pop Rocks"
[51] "Red vines"                    "Reese's Miniatures"
[53] "Reese's Peanut Butter cup"    "Reese's pieces"
[55] "Reese's stuffed with pieces" "Ring pop"
[57] "Rolo"                         "Root Beer Barrels"
[59] "Runts"                        "Sixlets"
[61] "Skittles original"            "Skittles wildberry"
[63] "Nestle Smarties"              "Smarties candy"
[65] "Snickers"                     "Snickers Crisper"
[67] "Sour Patch Kids"              "Sour Patch Tricksters"
[69] "Starburst"                    "Strawberry bon bons"
[71] "Sugar Babies"                 "Sugar Daddy"
[73] "Super Bubble"                 "Swedish Fish"
[75] "Tootsie Pop"                  "Tootsie Roll Juniors"
[77] "Tootsie Roll Midgies"         "Tootsie Roll Snack Bars"
[79] "Trolli Sour Bites"            "Twix"
[81] "Twizzlers"                    "Warheads"
[83] "Welch's Fruit Snacks"         "Werther's Original Caramel"
[85] "Whoppers"
```

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

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

```
[1] 38
```

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

```
[1] 37
```

What is your favorate candy?

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

```
candy["Milky Way", ]$winpercent
```

```
[1] 73.09956
```

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
```

```
library(dplyr)
```

```
Attaching package: 'dplyr'

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

    filter, lag

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

    intersect, setdiff, setequal, union
```

```
candy |>
  filter(rownames(candy) == "Haribo Happy Cola") |>
  select(winpercent)
```

```
                  winpercent
Haribo Happy Cola   34.15896
```

Q. Find fruity candy with a win percent above 50%

Using the filter feature:

```
candy |>
  filter(winpercent > 50) |>
  filter(fruity == 1)
```

```
                           chocolate fruity caramel peanutyalmondy nougat
Air Heads                          0      1       0              0      0
Haribo Gold Bears                  0      1       0              0      0
Haribo Sour Bears                  0      1       0              0      0
Lifesavers big ring gummies        0      1       0              0      0
Nerds                              0      1       0              0      0
Skittles original                  0      1       0              0      0
Skittles wildberry                 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
Swedish Fish                       0      1       0              0      0
                           crispedricewafer hard bar pluribus sugarpercent
Air Heads                                 0    0   0        0        0.906
Haribo Gold Bears                         0    0   0        1        0.465
Haribo Sour Bears                         0    0   0        1        0.465
Lifesavers big ring gummies               0    0   0        0        0.267
Nerds                                     0    1   0        1        0.848
Skittles original                         0    0   0        1        0.941
Skittles wildberry                        0    0   0        1        0.941
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
Swedish Fish                              0    0   0        1        0.604
                           pricepercent winpercent
Air Heads                         0.511   52.34146
Haribo Gold Bears                 0.465   57.11974
Haribo Sour Bears                 0.465   51.41243
Lifesavers big ring gummies       0.279   52.91139
Nerds                             0.325   55.35405
Skittles original                 0.220   63.08514
Skittles wildberry                0.220   55.10370
Sour Patch Kids                   0.116   59.86400
Sour Patch Tricksters             0.116   52.82595
```

```
Starburst                       0.220   67.03763
Swedish Fish                    0.755   54.86111
```

```
rownames(candy |>
  filter(winpercent > 50) |>
  filter(fruity == 1))
```

```
 [1] "Air Heads"                 "Haribo Gold Bears"
 [3] "Haribo Sour Bears"         "Lifesavers big ring gummies"
 [5] "Nerds"                     "Skittles original"
 [7] "Skittles wildberry"        "Sour Patch Kids"
 [9] "Sour Patch Tricksters"     "Starburst"
[11] "Swedish Fish"
```

Another way:

```
top.candy <- candy[candy$winpercent > 50, ]
top.candy[top.candy$fruity == 1, ]
```

```
                          chocolate fruity caramel peanutyalmondy nougat
Air Heads                         0      1       0              0      0
Haribo Gold Bears                 0      1       0              0      0
Haribo Sour Bears                 0      1       0              0      0
Lifesavers big ring gummies       0      1       0              0      0
Nerds                             0      1       0              0      0
Skittles original                 0      1       0              0      0
Skittles wildberry                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
Swedish Fish                      0      1       0              0      0
                          crispedricewafer hard bar pluribus sugarpercent
Air Heads                                 0    0   0        0        0.906
Haribo Gold Bears                         0    0   0        1        0.465
Haribo Sour Bears                         0    0   0        1        0.465
Lifesavers big ring gummies               0    0   0        0        0.267
Nerds                                     0    1   0        1        0.848
Skittles original                         0    0   0        1        0.941
Skittles wildberry                        0    0   0        1        0.941
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
Swedish Fish                             0    0   0        1        0.604
                           pricepercent winpercent
Air Heads                         0.511   52.34146
Haribo Gold Bears                 0.465   57.11974
Haribo Sour Bears                 0.465   51.41243
Lifesavers big ring gummies       0.279   52.91139
Nerds                             0.325   55.35405
Skittles original                 0.220   63.08514
Skittles wildberry                0.220   55.10370
Sour Patch Kids                   0.116   59.86400
Sour Patch Tricksters             0.116   52.82595
Starburst                         0.220   67.03763
Swedish Fish                      0.755   54.86111
```

```
rownames(top.candy[top.candy$fruity == 1, ])
```

```
 [1] "Air Heads"                  "Haribo Gold Bears"
 [3] "Haribo Sour Bears"          "Lifesavers big ring gummies"
 [5] "Nerds"                      "Skittles original"
 [7] "Skittles wildberry"         "Sour Patch Kids"
 [9] "Sour Patch Tricksters"      "Starburst"
[11] "Swedish Fish"
```

Usng skim to read the data:

```
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 `winpercent` column is measured differently than the rest of the other variables.

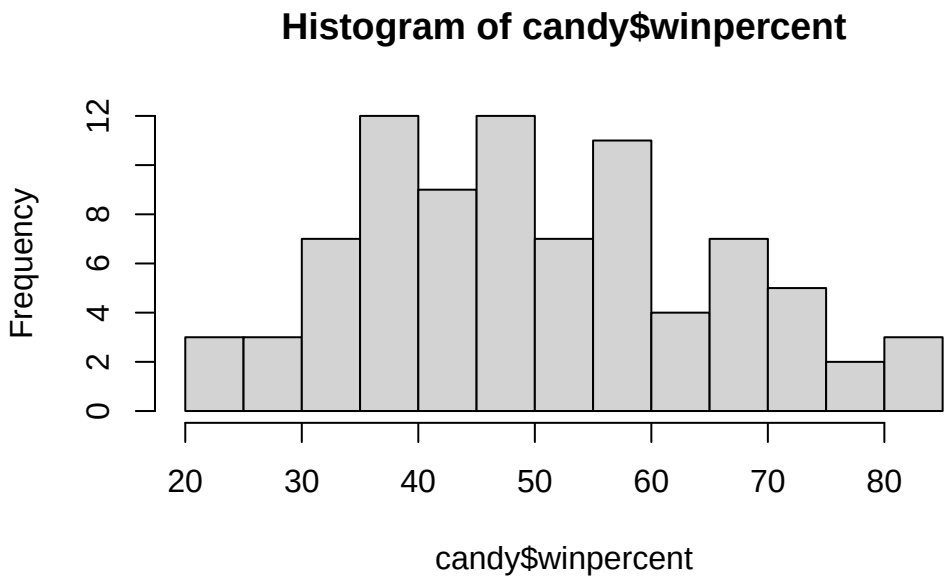
Q7. What do you think a zero and one represent for the candy$chocolate column?

The represent a "true" (1) or "false" (0) value.

Q8. Plot a histogram of winpercent values

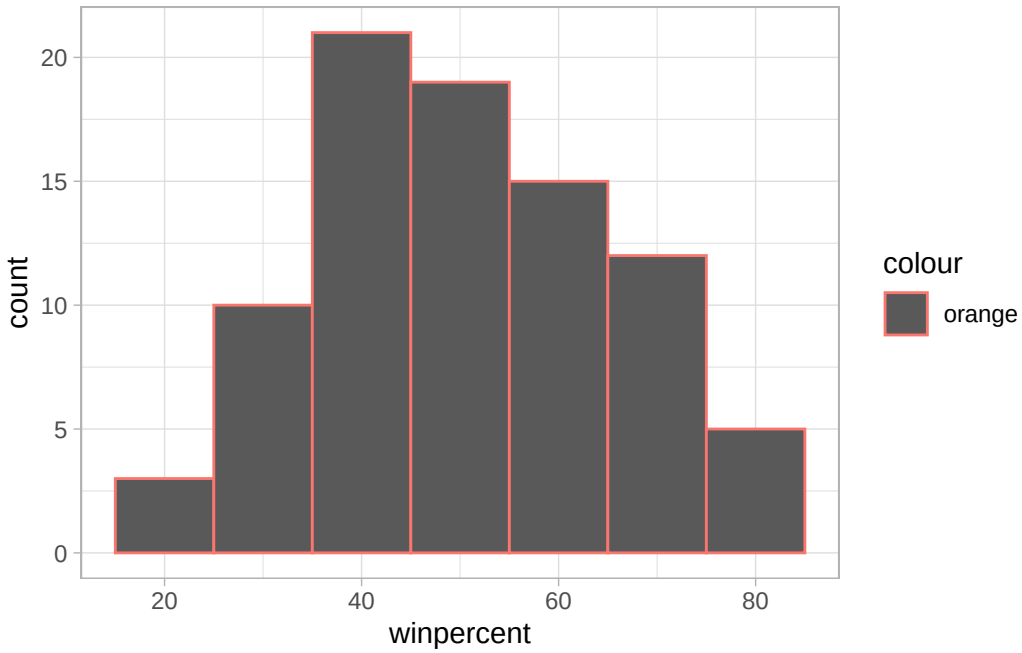
Using "base" R `hist()` function:

```r
hist(candy$winpercent, breaks = 20)
```



Using ggplot:

```r
library(ggplot2)
ggplot(candy, aes(x = winpercent, col = "orange")) +
geom_histogram(binwidth = 10) +
  theme_light()
```



Q9. Is the distribution of winpercent values symmetrical?

No

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

```
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?

```
fruit.candy <- candy |>
  filter(fruity == 1)

summary(fruit.candy$winpercent)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  22.45   39.04   42.97   44.12   52.11   67.04 
```

```
chocolate.candy <- candy |>
  filter(chocolate == 1)

summary(chocolate.candy$winpercent)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  34.72   50.35   60.80   60.92   70.74   84.18
```

Q12. Is this difference statistically significant?

```
t.test(chocolate.candy$winpercent, fruit.candy$winpercent)
```

```
    Welch Two Sample t-test

data:  chocolate.candy$winpercent and fruit.candy$winpercent
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
```

Small p-value means there is a statistically significant difference between fruity and chocolate candy.

Overall Candy Rankings

Learning to sort and order:

```
play <- c("d", "a", "c")
sort(play)
```

```
[1] "a" "c" "d"
```

```
order(play)
```

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

```
play[order(play)]
```

```
[1] "a" "c" "d"
```

```
sort(c(5,2,10), decreasing = T)
```

```
[1] 10  5  2
```

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

```
head(candy[order(candy$winpercent),], 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
```

```
rownames(head(candy[order(candy$winpercent),], 5))
```

```
[1] "Nik L Nip"          "Boston Baked Beans" "Chiclets"
[4] "Super Bubble"       "Jawbusters"
```

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

```
head(candy[order(candy$winpercent, decreasing = T),], 5)
```

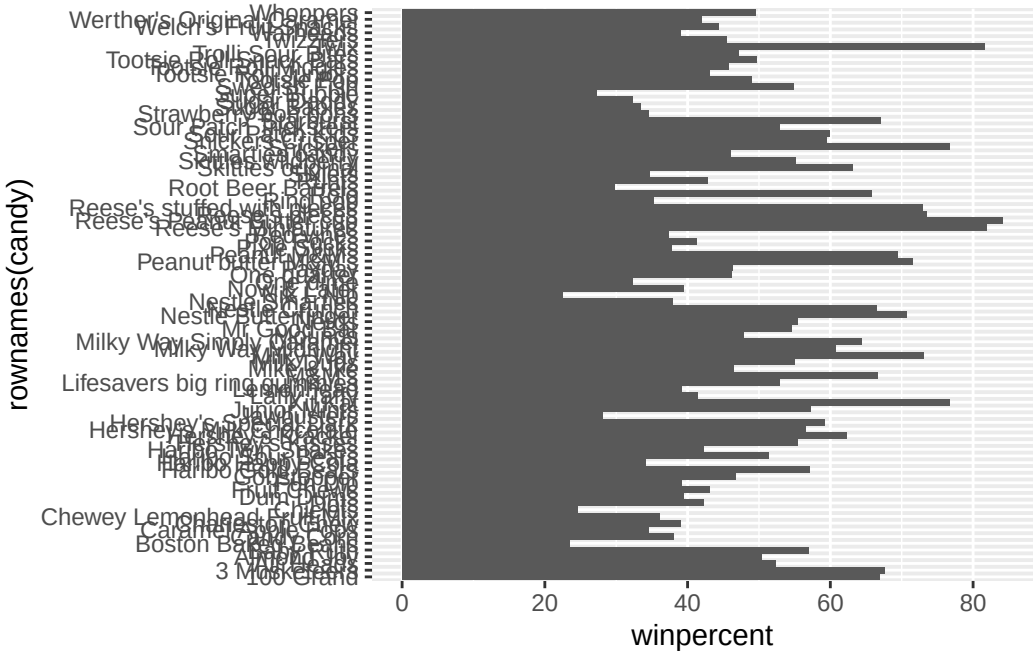
```
                         chocolate fruity caramel peanutyalmondy nougat
Reese's Peanut Butter cup         1      0       0              1      0
Reese's Miniatures                1      0       0              1      0
Twix                              1      0       1              0      0
Kit Kat                           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
```

```
rownames(head(candy[order(candy$winpercent, decreasing = T),], 5))
```

```
[1] "Reese's Peanut Butter cup" "Reese's Miniatures"
[3] "Twix"                      "Kit Kat"
[5] "Snickers"
```

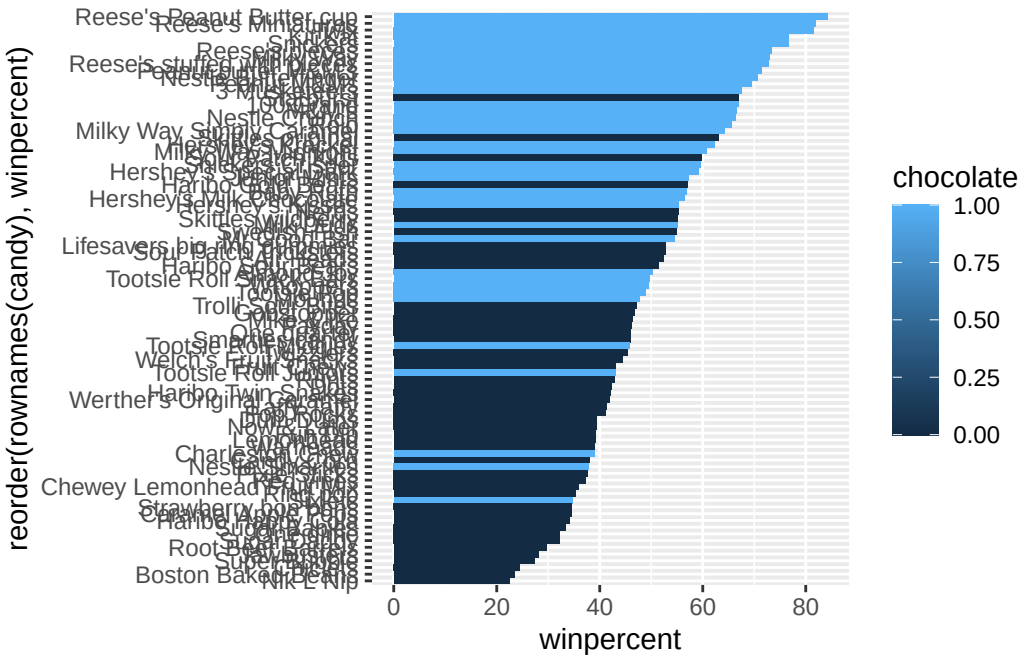
Q15. Make a first barplot of candy ranking based on winpercent values.

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



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

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



There needs to be a better color scheme to show the chocolate, fruity, caramel, etc categories. Using custom vector:

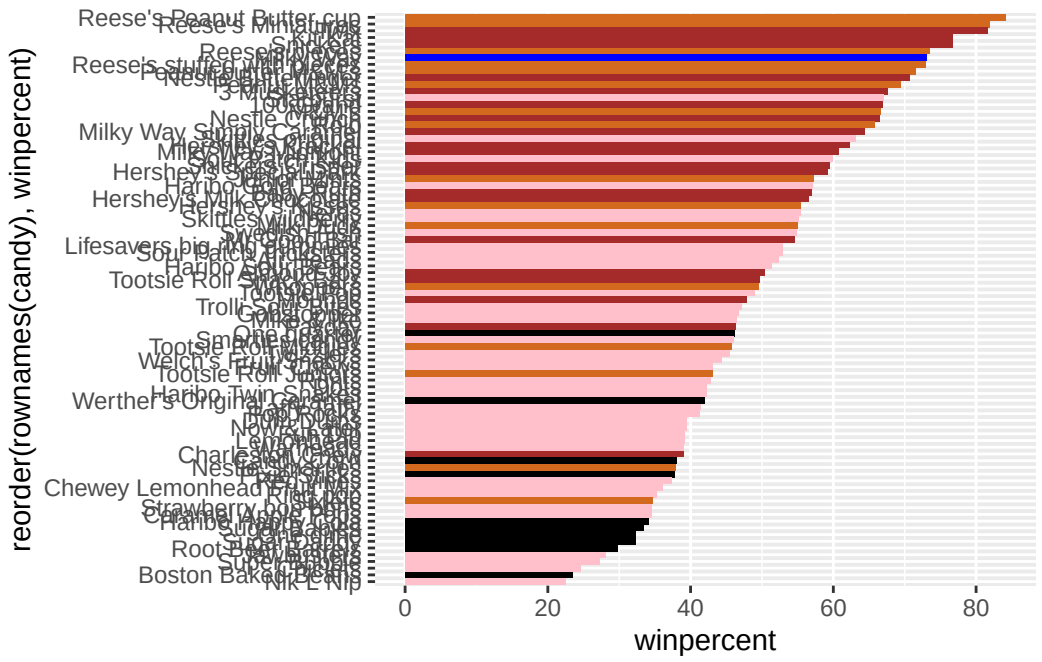
```
# Placr holder color vector

mycols <- rep("black", nrow(candy))
mycols[as.logical(candy$chocolate)] <- "chocolate"
mycols[as.logical(candy$bar)] <- "brown"
mycols[as.logical(candy$fruity)] <- "pink"

rownames(candy) == "Milky Way"
```

```
 [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[37]  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[85] FALSE
```

```
mycols[rownames(candy) == "Milky Way"] <- "blue"
mycols
```

```
 [1] "brown"     "brown"     "black"     "black"     "pink"      "brown"
 [7] "brown"     "black"     "black"     "pink"      "brown"     "pink"
[13] "pink"      "pink"      "pink"      "pink"      "pink"      "pink"
[19] "pink"      "black"     "pink"      "pink"      "chocolate" "brown"
[25] "brown"     "brown"     "pink"      "chocolate" "brown"     "pink"
[31] "pink"      "pink"      "chocolate" "chocolate" "pink"      "chocolate"
[37] "blue"      "brown"     "brown"     "brown"     "brown"     "pink"
[43] "brown"     "brown"     "pink"      "pink"      "brown"     "chocolate"
[49] "black"     "pink"      "pink"      "chocolate" "chocolate" "chocolate"
[55] "chocolate" "pink"      "chocolate" "black"     "pink"      "chocolate"
[61] "pink"      "pink"      "chocolate" "pink"      "brown"     "brown"
[67] "pink"      "pink"      "pink"      "pink"      "black"     "black"
[73] "pink"      "pink"      "pink"      "chocolate" "chocolate" "brown"
[79] "pink"      "brown"     "pink"      "pink"      "pink"      "black"
[85] "chocolate"
```

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



Q17. What is the worst ranked chocolate candy?

Sixlet

Q18. What is the best ranked fruity candy?

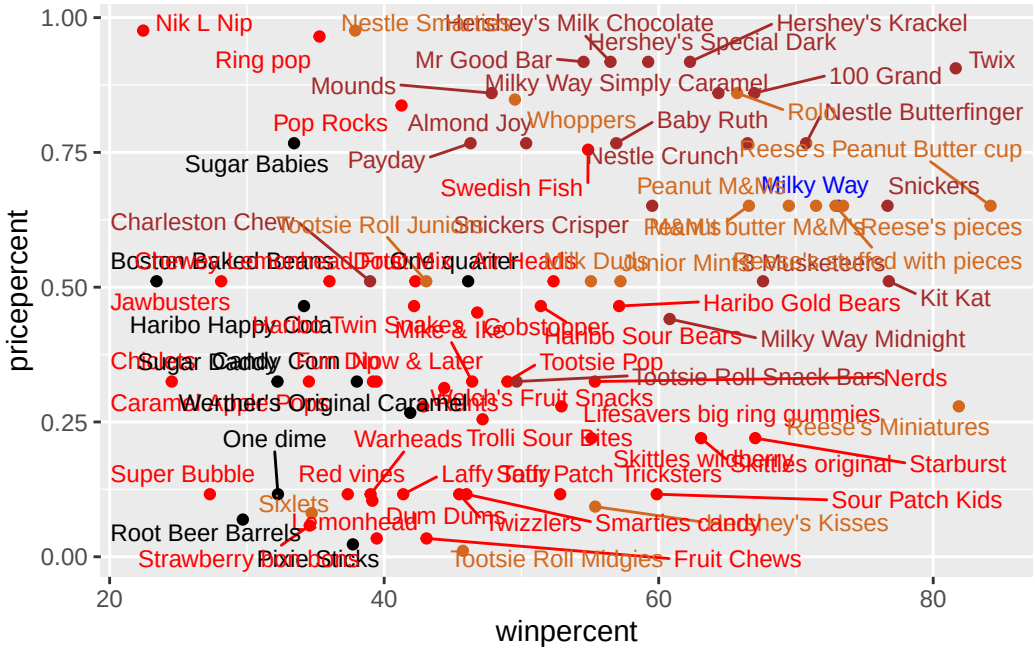
Starburst

Taking a look at pricepercent

Plotting winpercent vs pricepercent to see the best candy to buy:

```
mycols[as.logical(candy$fruity)] <- "red"
```

```
library(ggrepel)

# How about a plot of price vs win
ggplot(candy) +
  aes(winpercent, pricepercent, label=rownames(candy)) +
  geom_point(col=mycols) +
  geom_text_repel(col=mycols, size=3.3, max.overlaps = 30)
```



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

Reese's miniatures

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

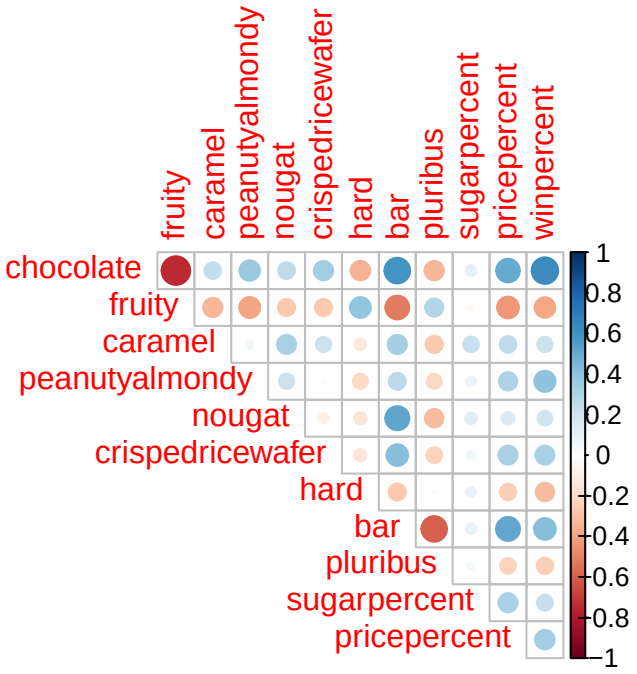
Nik L Lip, Ring Pop, Nestle Smarties, Mr. Goodbar, and Milky Way. Nik L Lip is the least popular.

Exploring the correlation structure

```
library(corrplot)
```

```
corrplot 0.95 loaded
```

```
cij <- cor(candy)
corrplot(cij, diag = F, type = "upper")
```



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

Chocolate/fruity and plurbius/bar

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

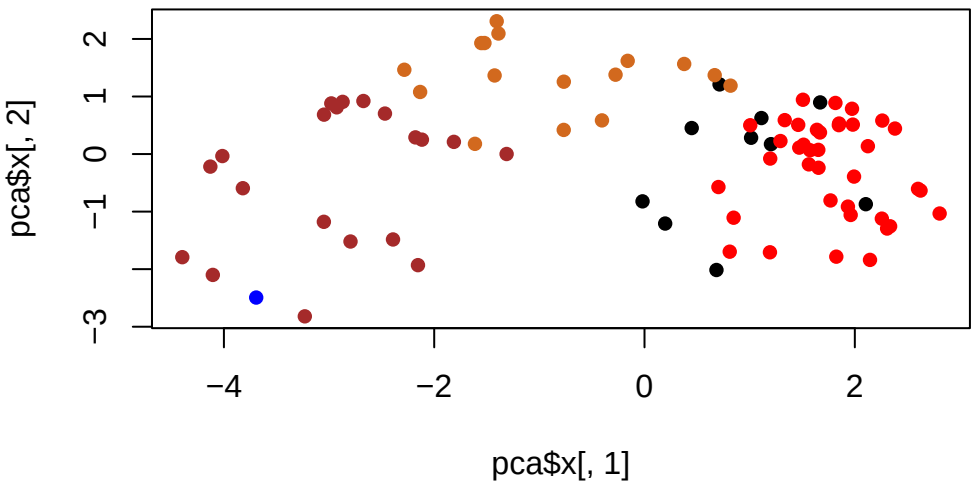
winpercent/chocolate and bar/chocolate

Principal Component Analysis

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

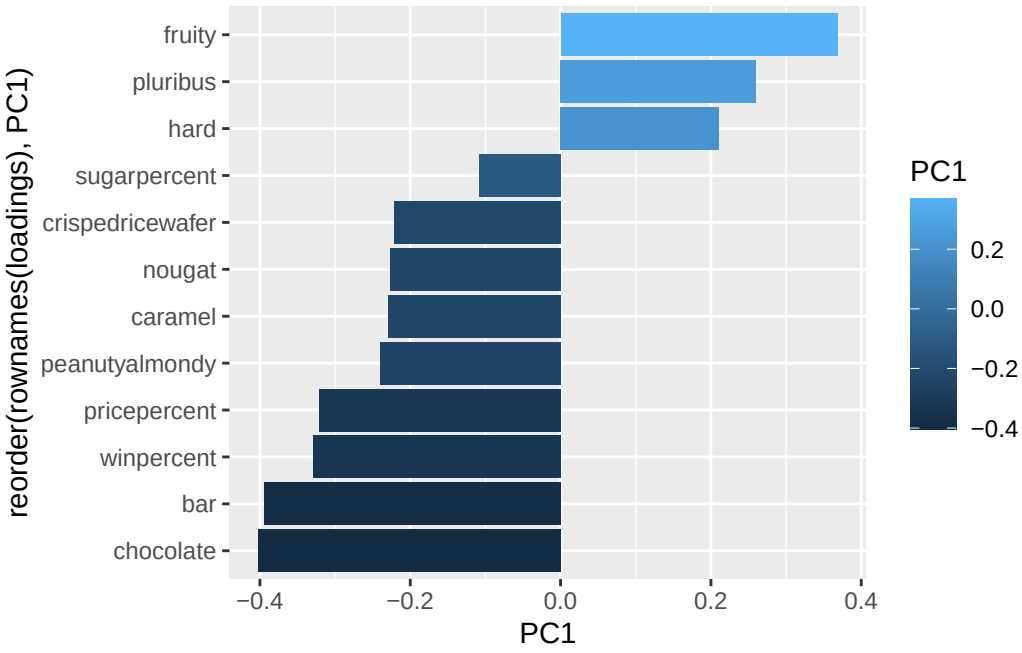
```
Importance of components:
                          PC1    PC2    PC3     PC4    PC5     PC6     PC7
Standard deviation     2.0788 1.1378 1.1092 1.07533 0.9518 0.81923 0.81530
Proportion of Variance 0.3601 0.1079 0.1025 0.09636 0.0755 0.05593 0.05539
Cumulative Proportion  0.3601 0.4680 0.5705 0.66688 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
```

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

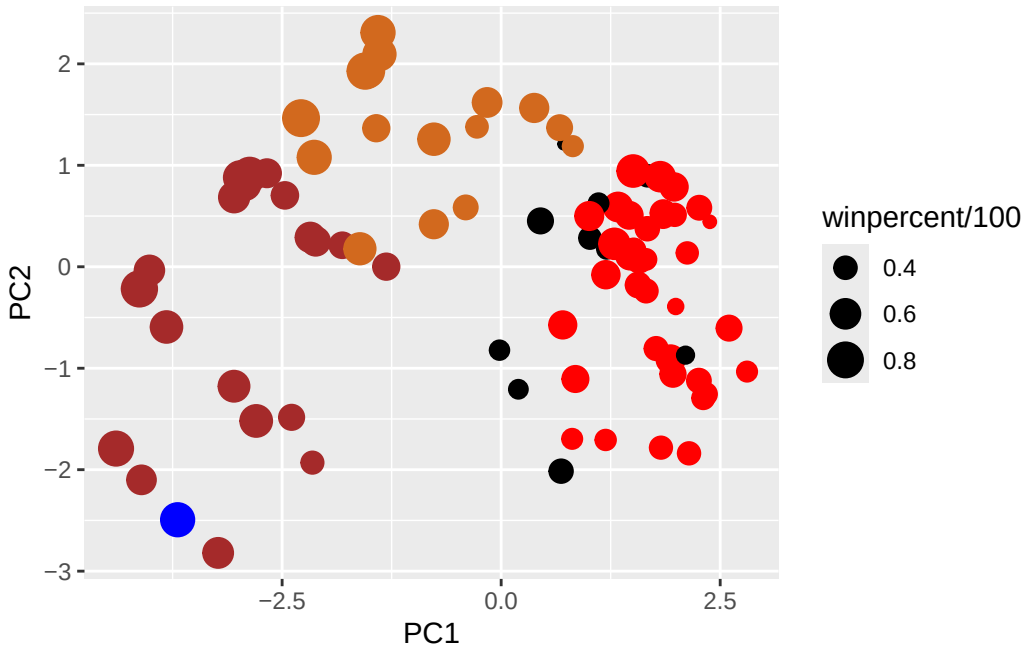


Showing how the original variables contribute to the new PCs. PC1:

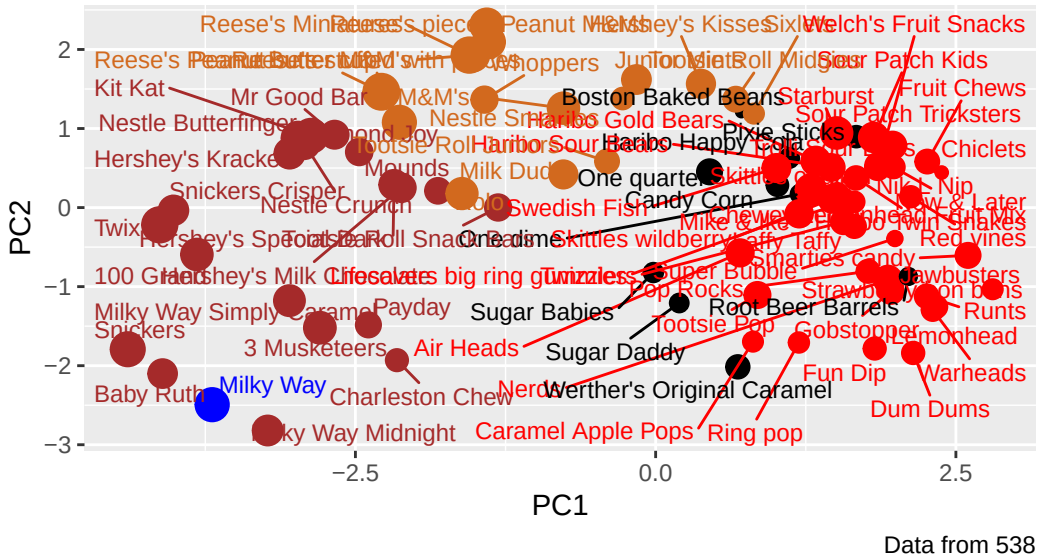
```
loadings <- as.data.frame(pca$rotation)

ggplot(loadings, aes(x = PC1, y = reorder(rownames(loadings), PC1), fill = PC1)) +
  geom_col()
```




```
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
```

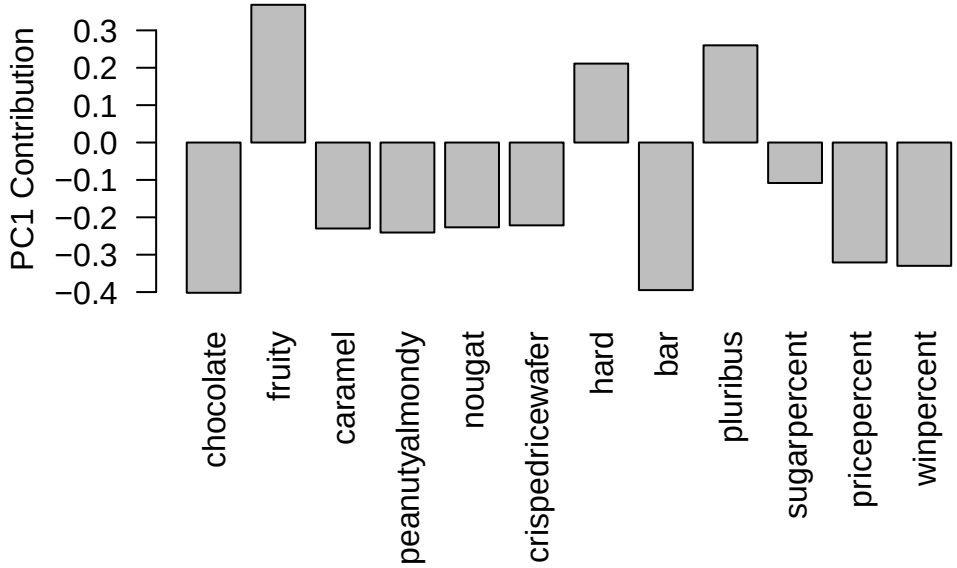



```
library(ggrepel)

p + geom_text_repel(size=3.3, col=mycols, max.overlaps = 40)  +
  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")
```




```
#library(plotly)
#ggplotly(p)
```

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



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

Fruity, hard, and plurbius