Lab 10 Halloween Analysis

AUTHOR
Shane Desfor

Lab 10 Halloween Data analysis

```
candy_file <- "candy-data.csv"
##Set Row Names to First Column:

candy = read.csv("candy-data.csv", row.names=1)
sum(candy$fruity)
```

```
[1] 38
```

> Q1. How many different candy types are in this dataset? Ans. nrow(candy), 85 Q2. How many fruity candy types are in the dataset? Ans., 38 The functions dim(), nrow(), table() and sum() may be useful for answering the first 2 questions.

```
nrow(candy)
```

```
[1] 85
```

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

```
[1] 38
```

> Q3. What is your favorite candy in the dataset and what is it's winpercent value? Can Call row.names to see all the candies. Then call win percent of that candy. Ans.84% Q4. What is the winpercent value for "Kit Kat"? Ans. 77% Q5. What is the winpercent value for "Tootsie Roll Snack Bars"? Ans. 50%

```
row.names(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] "HersheyÕs Kisses"           "HersheyÕs Krackel"
[25] "HersheyÕs Milk Chocolate"   "HersheyÕs Special Dark"
[27] "Jawbusters"                 "Junior Mints"
[29] "Kit Kat"                    "Laffy Taffy"
[31] "Lemonhead"                  "Lifesavers big ring gummies"
[33] "Peanut butter M&MÕs"        "M&MÕs"
[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"                  "ReeseÕs Miniatures"
[53] "ReeseÕs Peanut Butter cup"  "ReeseÕs pieces"
[55] "ReeseÕs 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] "WelchÕs Fruit Snacks"       "WertherÕs Original Caramel"
[85] "Whoppers"
```

```
candy[ "ReeseÕs Peanut Butter cup",  ]
```

```
                         chocolate fruity caramel peanutyalmondy nougat
ReeseÕs Peanut Butter cup         1      0       0              1      0
                         crispedricewafer hard bar pluribus sugarpercent
ReeseÕs Peanut Butter cup                 0    0   0        0         0.72
                         pricepercent winpercent
ReeseÕs Peanut Butter cup        0.651   84.18029
```

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

```
        chocolate fruity caramel peanutyalmondy nougat crispedricewafer hard
Kit Kat         1      0       0              0      0                 1    0
        bar pluribus sugarpercent pricepercent winpercent
Kit Kat   1        0        0.313        0.511    76.7686
```

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

```
                        chocolate fruity caramel peanutyalmondy nougat
Tootsie Roll Snack Bars         1      0       0              0      0
                        crispedricewafer hard bar pluribus sugarpercent
Tootsie Roll Snack Bars                0    0   1        0        0.465
                        pricepercent winpercent
Tootsie Roll Snack Bars        0.325    49.6535
```

#Can pull a function from a package instead of moving through library

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

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

Data summary

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

Variable type: numeric

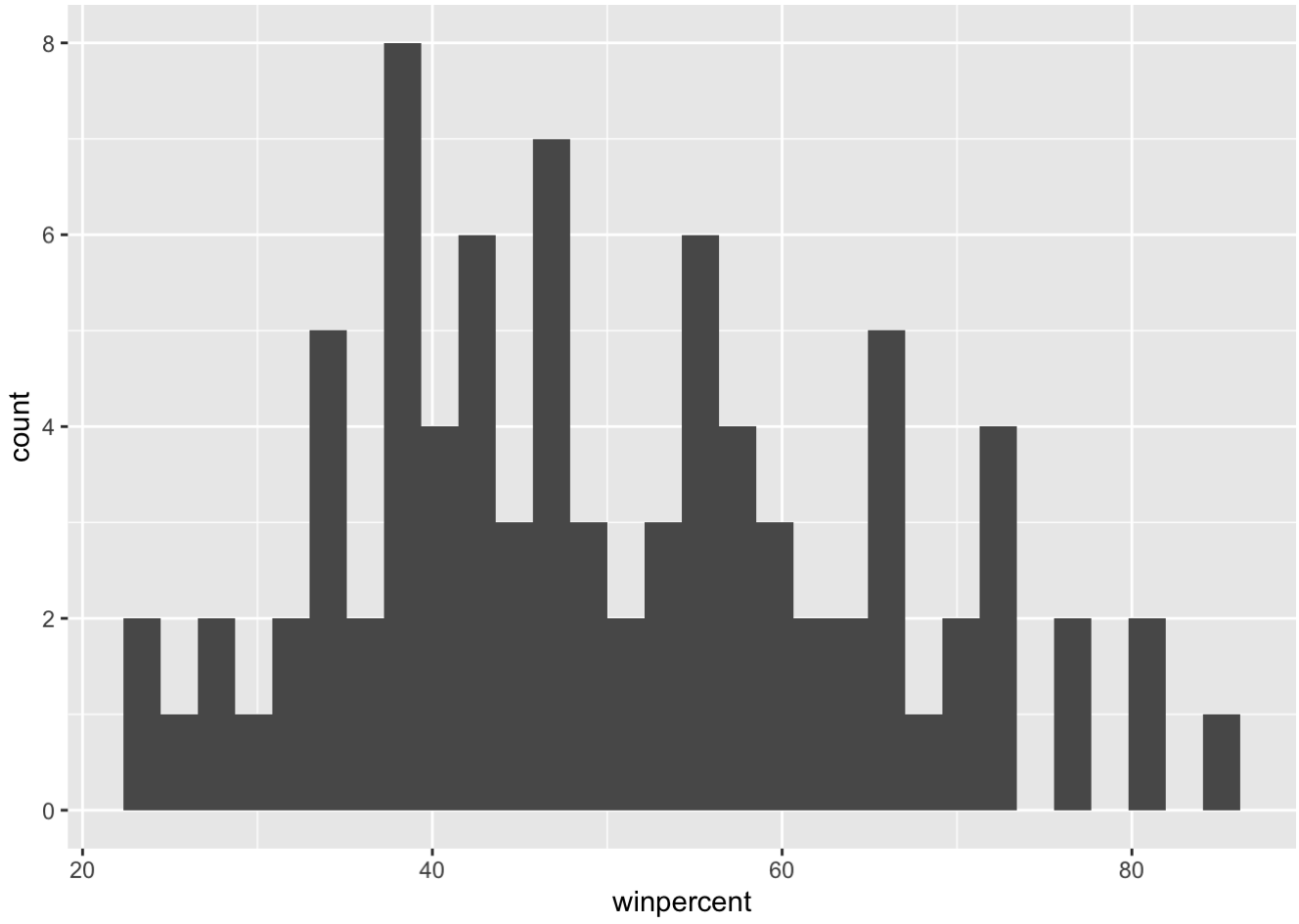
skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
chocolate	0	1	0.44	0.50	0.00	0.00	0.00	1.00	1.00	▇▁▁▁▅
fruity	0	1	0.45	0.50	0.00	0.00	0.00	1.00	1.00	▇▁▁▁▅
caramel	0	1	0.16	0.37	0.00	0.00	0.00	0.00	1.00	▇▁▁▁▂
peanutyalmondy	0	1	0.16	0.37	0.00	0.00	0.00	0.00	1.00	▇▁▁▁▂
nougat	0	1	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	▇▁▁▁▂

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
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	

> Q8. Plot a histogram of winpercent values Q9. Is the distribution of winpercent values symmetrical? Ans. No Q10. Is the center of the distribution above or below 50%? Ans. Below

```
library("ggplot2")

ggplot(candy, aes(winpercent, bine=10,)) + geom_histogram()
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
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

> Q11. On average is chocolate candy higher or lower ranked than fruit candy? chocolate is 61%, fruity is 44% Q12. Is this difference statistically significant? Yes, p-value = 2.871e-08

```
chocolate.ind <- as.logical(candy$chocolate)
candy[ chocolate.ind, ]
```

	chocolate	fruity	caramel	peanutyalmondy	nougat
100 Grand	1	0	1	0	0
3 Musketeers	1	0	0	0	1
Almond Joy	1	0	0	1	0
Baby Ruth	1	0	1	1	1
Charleston Chew	1	0	0	0	1
HersheyÕs Kisses	1	0	0	0	0
HersheyÕs Krackel	1	0	0	0	0
HersheyÕs Milk Chocolate	1	0	0	0	0
HersheyÕs Special Dark	1	0	0	0	0
Junior Mints	1	0	0	0	0
Kit Kat	1	0	0	0	0
Peanut butter M&MÕs	1	0	0	1	0
M&MÕs	1	0	0	0	0
Milk Duds	1	0	1	0	0
Milky Way	1	0	1	0	1
Milky Way Midnight	1	0	1	0	1
Milky Way Simply Caramel	1	0	1	0	0
Mounds	1	0	0	0	0
Mr Good Bar	1	0	0	1	0
Nestle Butterfinger	1	0	0	1	0
Nestle Crunch	1	0	0	0	0
Peanut M&Ms	1	0	0	1	0
ReeseÕs Miniatures	1	0	0	1	0
ReeseÕs Peanut Butter cup	1	0	0	1	0
ReeseÕs pieces	1	0	0	1	0
ReeseÕs stuffed with pieces	1	0	0	1	0
Rolo	1	0	1	0	0
Sixlets	1	0	0	0	0
Nestle Smarties	1	0	0	0	0

Nestle Smarties	1	0	0	0	0
Snickers	1	0	1	1	1
Snickers Crisper	1	0	1	1	0
Tootsie Pop	1	1	0	0	0
Tootsie Roll Juniors	1	0	0	0	0
Tootsie Roll Midgies	1	0	0	0	0
Tootsie Roll Snack Bars	1	0	0	0	0
Twix	1	0	1	0	0
Whoppers	1	0	0	0	0

	crisped	ricewafer	hard	bar	pluribus	sugarpercent
100 Grand	1	0	1	0		0.732
3 Musketeers	0	0	1	0		0.604
Almond Joy	0	0	1	0		0.465
Baby Ruth	0	0	1	0		0.604
Charleston Chew	0	0	1	0		0.604
HersheyÕs Kisses	0	0	0	1		0.127
HersheyÕs Krackel	1	0	1	0		0.430
HersheyÕs Milk Chocolate	0	0	1	0		0.430
HersheyÕs Special Dark	0	0	1	0		0.430
Junior Mints	0	0	0	1		0.197
Kit Kat	1	0	1	0		0.313
Peanut butter M&MÕs	0	0	0	1		0.825
M&MÕs	0	0	0	1		0.825
Milk Duds	0	0	0	1		0.302
Milky Way	0	0	1	0		0.604
Milky Way Midnight	0	0	1	0		0.732
Milky Way Simply Caramel	0	0	1	0		0.965
Mounds	0	0	1	0		0.313
Mr Good Bar	0	0	1	0		0.313
Nestle Butterfinger	0	0	1	0		0.604
Nestle Crunch	1	0	1	0		0.313
Peanut M&Ms	0	0	0	1		0.593
ReeseÕs Miniatures	0	0	0	0		0.034
ReeseÕs Peanut Butter cup	0	0	0	0		0.720
ReeseÕs pieces	0	0	0	1		0.406
ReeseÕs stuffed with pieces	0	0	0	0		0.988
Rolo	0	0	0	1		0.860
Sixlets	0	0	0	1		0.220
Nestle Smarties	0	0	0	1		0.267
Snickers	0	0	1	0		0.546
Snickers Crisper	1	0	1	0		0.604
Tootsie Pop	0	1	0	0		0.604
Tootsie Roll Juniors	0	0	0	0		0.313
Tootsie Roll Midgies	0	0	0	1		0.174
Tootsie Roll Snack Bars	0	0	1	0		0.465
Twix	1	0	1	0		0.546
Whoppers	1	0	0	1		0.872

	pricepercent	winpercent
100 Grand	0.860	66.97173
3 Musketeers	0.511	67.60294
Almond Joy	0.767	50.34755

Baby Ruth	0.767	56.91455
Charleston Chew	0.511	38.97504
HersheyÕs Kisses	0.093	55.37545
HersheyÕs Krackel	0.918	62.28448
HersheyÕs Milk Chocolate	0.918	56.49050
HersheyÕs Special Dark	0.918	59.23612
Junior Mints	0.511	57.21925
Kit Kat	0.511	76.76860
Peanut butter M&MÕs	0.651	71.46505
M&MÕs	0.651	66.57458
Milk Duds	0.511	55.06407
Milky Way	0.651	73.09956
Milky Way Midnight	0.441	60.80070
Milky Way Simply Caramel	0.860	64.35334
Mounds	0.860	47.82975
Mr Good Bar	0.918	54.52645
Nestle Butterfinger	0.767	70.73564
Nestle Crunch	0.767	66.47068
Peanut M&Ms	0.651	69.48379
ReeseÕs Miniatures	0.279	81.86626
ReeseÕs Peanut Butter cup	0.651	84.18029
ReeseÕs pieces	0.651	73.43499
ReeseÕs stuffed with pieces	0.651	72.88790
Rolo	0.860	65.71629
Sixlets	0.081	34.72200
Nestle Smarties	0.976	37.88719
Snickers	0.651	76.67378
Snickers Crisper	0.651	59.52925
Tootsie Pop	0.325	48.98265
Tootsie Roll Juniors	0.511	43.06890
Tootsie Roll Midgies	0.011	45.73675
Tootsie Roll Snack Bars	0.325	49.65350
Twix	0.906	81.64291
Whoppers	0.848	49.52411

```
##This is how you call winpercent OF chocolate logical

chocolate.win <- candy[ chocolate.ind, ]$winpercent

##Then to summarize, call mean
mean(chocolate.win)
```

```
[1] 60.92153
```

```
##Now fruity
fruity.ind <- as.logical((candy$fruity))
fruity.win <- candy[fruity.ind, ]$winpercent
mean(fruity.win)
```

```
[1] 44.11974
```

```
t.test(fruity.win, chocolate.win)
```

```
    Welch Two Sample t-test

data:  fruity.win and chocolate.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:
 -22.15795 -11.44563
sample estimates:
mean of x mean of y
 44.11974  60.92153
```

##3. Overal Candy Rankings can use the sort function and order function

```
x <- c(5, 6, 1, 9, 10)
sort(x)
```

```
[1]  1  5  6  9 10
```

order shows the order of the vector, where each value shows the place the element should be in that order but put it as a vector of location and order, get them ordered.

```
order(x)
```

```
[1] 3 1 2 4 5
```

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

```
[1]  1  5  6  9 10
```

> Q13. What are the five least liked candy types in this set? Nik Lnip, Boston Baked Beans, Chiclets, Super Bubble, Jaw Busters Q14. What are the top 5 all time favorite candy types out of this set?

```
order(candy$winpercent)
```

```
 [1] 45  8 13 73 27 58 72  3 71 20 10 70 60 56 12 51 49 63  9 11 82 31 17 46 15
[26] 50 30 84 22 14 59 76 16 83 81 77 64  4 47 35 18 79 40 75 85 78  6 21  5 68
[51] 32 41 74 36 62 42 23 25  7 19 28 26 66 67 38 24 61 39 57 44 34  1 69  2 48
[76] 43 33 55 37 54 65 29 80 52 53
```

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

```
                         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
Sugar Daddy	0	0	1	0	0
One dime	0	0	0	0	0
Sugar Babies	0	0	1	0	0
Haribo Happy Cola	0	0	0	0	0
Caramel Apple Pops	0	1	1	0	0
Strawberry bon bons	0	1	0	0	0
Sixlets	1	0	0	0	0
Ring pop	0	1	0	0	0
Chewey Lemonhead Fruit Mix	0	1	0	0	0
Red vines	0	1	0	0	0
Pixie Sticks	0	0	0	0	0
Nestle Smarties	1	0	0	0	0
Candy Corn	0	0	0	0	0
Charleston Chew	1	0	0	0	1
Warheads	0	1	0	0	0
Lemonhead	0	1	0	0	0
Fun Dip	0	1	0	0	0
Now & Later	0	1	0	0	0
Dum Dums	0	1	0	0	0
Pop Rocks	0	1	0	0	0
Laffy Taffy	0	1	0	0	0
WertherÕs Original Caramel	0	0	1	0	0
Haribo Twin Snakes	0	1	0	0	0
Dots	0	1	0	0	0
Runts	0	1	0	0	0
Tootsie Roll Juniors	1	0	0	0	0
Fruit Chews	0	1	0	0	0
WelchÕs Fruit Snacks	0	1	0	0	0
Twizzlers	0	1	0	0	0
Tootsie Roll Midgies	1	0	0	0	0
Smarties candy	0	1	0	0	0
One quarter	0	0	0	0	0
Payday	0	0	0	1	1
Mike & Ike	0	1	0	0	0
Gobstopper	0	1	0	0	0
Trolli Sour Bites	0	1	0	0	0
Mounds	1	0	0	0	0
Tootsie Pop	1	1	0	0	0
Whoppers	1	0	0	0	0
Tootsie Roll Snack Bars	1	0	0	0	0
Almond Joy	1	0	0	1	0
Haribo Sour Bears	0	1	0	0	0
Air Heads	0	1	0	0	0
Sour Patch Tricksters	0	1	0	0	0
Lifesavers big ring gummies	0	1	0	0	0

Mr Good Bar	1	0	0	1	0
Swedish Fish	0	1	0	0	0
Milk Duds	1	0	1	0	0
Skittles wildberry	0	1	0	0	0
Nerds	0	1	0	0	0
HersheyÕs Kisses	1	0	0	0	0
HersheyÕs Milk Chocolate	1	0	0	0	0
Baby Ruth	1	0	1	1	1
Haribo Gold Bears	0	1	0	0	0
Junior Mints	1	0	0	0	0
HersheyÕs Special Dark	1	0	0	0	0
Snickers Crisper	1	0	1	1	0
Sour Patch Kids	0	1	0	0	0
Milky Way Midnight	1	0	1	0	1
HersheyÕs Krackel	1	0	0	0	0
Skittles original	0	1	0	0	0
Milky Way Simply Caramel	1	0	1	0	0
Rolo	1	0	1	0	0
Nestle Crunch	1	0	0	0	0
M&MÕs	1	0	0	0	0
100 Grand	1	0	1	0	0
Starburst	0	1	0	0	0
3 Musketeers	1	0	0	0	1
Peanut M&Ms	1	0	0	1	0
Nestle Butterfinger	1	0	0	1	0
Peanut butter M&MÕs	1	0	0	1	0
ReeseÕs stuffed with pieces	1	0	0	1	0
Milky Way	1	0	1	0	1
ReeseÕs pieces	1	0	0	1	0
Snickers	1	0	1	1	1
Kit Kat	1	0	0	0	0
Twix	1	0	1	0	0
ReeseÕs Miniatures	1	0	0	1	0
ReeseÕs Peanut Butter cup	1	0	0	1	0

	crispedricewafer	hard	bar	pluribus	sugarpercent
Nik L Nip	0	0	0	1	0.197
Boston Baked Beans	0	0	0	1	0.313
Chiclets	0	0	0	1	0.046
Super Bubble	0	0	0	0	0.162
Jawbusters	0	1	0	1	0.093
Root Beer Barrels	0	1	0	1	0.732
Sugar Daddy	0	0	0	0	0.418
One dime	0	0	0	0	0.011
Sugar Babies	0	0	0	1	0.965
Haribo Happy Cola	0	0	0	1	0.465
Caramel Apple Pops	0	0	0	0	0.604
Strawberry bon bons	0	1	0	1	0.569
Sixlets	0	0	0	1	0.220
Ring pop	0	1	0	0	0.732
Chewey Lemonhead Fruit Mix	0	0	0	1	0.732
Red vines	0	0	0	1	0.581

Pixie Sticks	0	0	0	1	0.093
Nestle Smarties	0	0	0	1	0.267
Candy Corn	0	0	0	1	0.906
Charleston Chew	0	0	1	0	0.604
Warheads	0	1	0	0	0.093
Lemonhead	0	1	0	0	0.046
Fun Dip	0	1	0	0	0.732
Now & Later	0	0	0	1	0.220
Dum Dums	0	1	0	0	0.732
Pop Rocks	0	1	0	1	0.604
Laffy Taffy	0	0	0	0	0.220
WertherÕs Original Caramel	0	1	0	0	0.186
Haribo Twin Snakes	0	0	0	1	0.465
Dots	0	0	0	1	0.732
Runts	0	1	0	1	0.872
Tootsie Roll Juniors	0	0	0	0	0.313
Fruit Chews	0	0	0	1	0.127
WelchÕs Fruit Snacks	0	0	0	1	0.313
Twizzlers	0	0	0	0	0.220
Tootsie Roll Midgies	0	0	0	1	0.174
Smarties candy	0	1	0	1	0.267
One quarter	0	0	0	0	0.011
Payday	0	0	1	0	0.465
Mike & Ike	0	0	0	1	0.872
Gobstopper	0	1	0	1	0.906
Trolli Sour Bites	0	0	0	1	0.313
Mounds	0	0	1	0	0.313
Tootsie Pop	0	1	0	0	0.604
Whoppers	1	0	0	1	0.872
Tootsie Roll Snack Bars	0	0	1	0	0.465
Almond Joy	0	0	1	0	0.465
Haribo Sour Bears	0	0	0	1	0.465
Air Heads	0	0	0	0	0.906
Sour Patch Tricksters	0	0	0	1	0.069
Lifesavers big ring gummies	0	0	0	0	0.267
Mr Good Bar	0	0	1	0	0.313
Swedish Fish	0	0	0	1	0.604
Milk Duds	0	0	0	1	0.302
Skittles wildberry	0	0	0	1	0.941
Nerds	0	1	0	1	0.848
HersheyÕs Kisses	0	0	0	1	0.127
HersheyÕs Milk Chocolate	0	0	1	0	0.430
Baby Ruth	0	0	1	0	0.604
Haribo Gold Bears	0	0	0	1	0.465
Junior Mints	0	0	0	1	0.197
HersheyÕs Special Dark	0	0	1	0	0.430
Snickers Crisper	1	0	1	0	0.604
Sour Patch Kids	0	0	0	1	0.069
Milky Way Midnight	0	0	1	0	0.732
HersheyÕs Krackel	1	0	1	0	0.430
Skittles original	0	0	0	1	0.941

Milky Way Simply Caramel	0	0	1	0	0.965
Rolo	0	0	0	1	0.860
Nestle Crunch	1	0	1	0	0.313
M&MÕs	0	0	0	1	0.825
100 Grand	1	0	1	0	0.732
Starburst	0	0	0	1	0.151
3 Musketeers	0	0	1	0	0.604
Peanut M&Ms	0	0	0	1	0.593
Nestle Butterfinger	0	0	1	0	0.604
Peanut butter M&MÕs	0	0	0	1	0.825
ReeseÕs stuffed with pieces	0	0	0	0	0.988
Milky Way	0	0	1	0	0.604
ReeseÕs pieces	0	0	0	1	0.406
Snickers	0	0	1	0	0.546
Kit Kat	1	0	1	0	0.313
Twix	1	0	1	0	0.546
ReeseÕs Miniatures	0	0	0	0	0.034
ReeseÕs Peanut Butter cup	0	0	0	0	0.720

	pricepercent	winpercent
Nik L Nip	0.976	22.44534
Boston Baked Beans	0.511	23.41782
Chiclets	0.325	24.52499
Super Bubble	0.116	27.30386
Jawbusters	0.511	28.12744
Root Beer Barrels	0.069	29.70369
Sugar Daddy	0.325	32.23100
One dime	0.116	32.26109
Sugar Babies	0.767	33.43755
Haribo Happy Cola	0.465	34.15896
Caramel Apple Pops	0.325	34.51768
Strawberry bon bons	0.058	34.57899
Sixlets	0.081	34.72200
Ring pop	0.965	35.29076
Chewey Lemonhead Fruit Mix	0.511	36.01763
Red vines	0.116	37.34852
Pixie Sticks	0.023	37.72234
Nestle Smarties	0.976	37.88719
Candy Corn	0.325	38.01096
Charleston Chew	0.511	38.97504
Warheads	0.116	39.01190
Lemonhead	0.104	39.14106
Fun Dip	0.325	39.18550
Now & Later	0.325	39.44680
Dum Dums	0.034	39.46056
Pop Rocks	0.837	41.26551
Laffy Taffy	0.116	41.38956
WertherÕs Original Caramel	0.267	41.90431
Haribo Twin Snakes	0.465	42.17877
Dots	0.511	42.27208
Runts	0.279	42.84914
Tootsie Roll Juniors	0.511	43.06890

Fruit Chews	0.034	43.08892
WelchÕs Fruit Snacks	0.313	44.37552
Twizzlers	0.116	45.46628
Tootsie Roll Midgies	0.011	45.73675
Smarties candy	0.116	45.99583
One quarter	0.511	46.11650
Payday	0.767	46.29660
Mike & Ike	0.325	46.41172
Gobstopper	0.453	46.78335
Trolli Sour Bites	0.255	47.17323
Mounds	0.860	47.82975
Tootsie Pop	0.325	48.98265
Whoppers	0.848	49.52411
Tootsie Roll Snack Bars	0.325	49.65350
Almond Joy	0.767	50.34755
Haribo Sour Bears	0.465	51.41243
Air Heads	0.511	52.34146
Sour Patch Tricksters	0.116	52.82595
Lifesavers big ring gummies	0.279	52.91139
Mr Good Bar	0.918	54.52645
Swedish Fish	0.755	54.86111
Milk Duds	0.511	55.06407
Skittles wildberry	0.220	55.10370
Nerds	0.325	55.35405
HersheyÕs Kisses	0.093	55.37545
HersheyÕs Milk Chocolate	0.918	56.49050
Baby Ruth	0.767	56.91455
Haribo Gold Bears	0.465	57.11974
Junior Mints	0.511	57.21925
HersheyÕs Special Dark	0.918	59.23612
Snickers Crisper	0.651	59.52925
Sour Patch Kids	0.116	59.86400
Milky Way Midnight	0.441	60.80070
HersheyÕs Krackel	0.918	62.28448
Skittles original	0.220	63.08514
Milky Way Simply Caramel	0.860	64.35334
Rolo	0.860	65.71629
Nestle Crunch	0.767	66.47068
M&MÕs	0.651	66.57458
100 Grand	0.860	66.97173
Starburst	0.220	67.03763
3 Musketeers	0.511	67.60294
Peanut M&Ms	0.651	69.48379
Nestle Butterfinger	0.767	70.73564
Peanut butter M&MÕs	0.651	71.46505
ReeseÕs stuffed with pieces	0.651	72.88790
Milky Way	0.651	73.09956
ReeseÕs pieces	0.651	73.43499
Snickers	0.651	76.67378
Kit Kat	0.511	76.76860
Twix	0.906	81.64291

```
ReeseÕs Miniatures                0.279   81.86626
ReeseÕs Peanut Butter cup         0.651   84.18029
```

```
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
```

```
#Tidy vers version
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
```

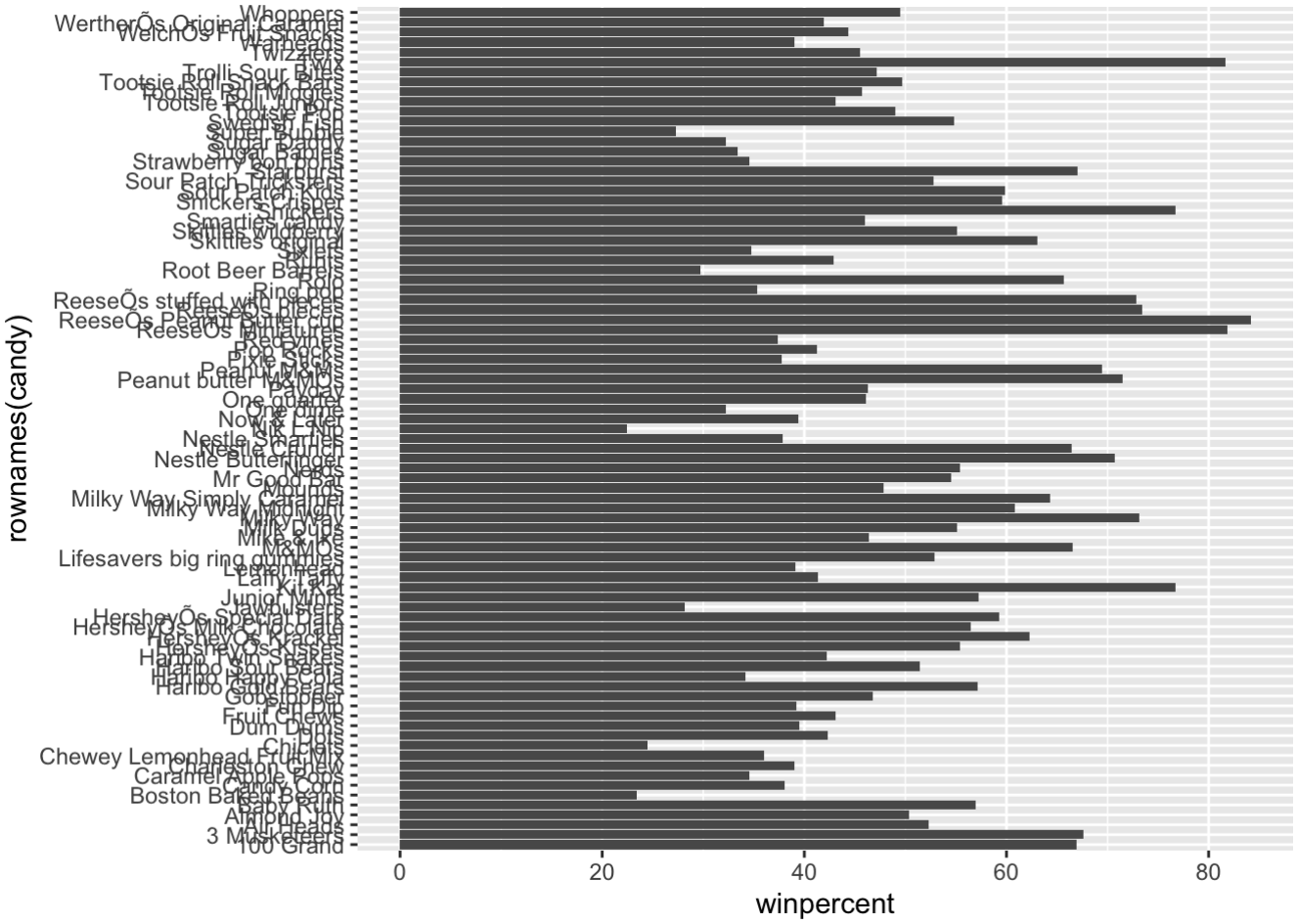
```
candy %>% arrange(winpercent) %>% head(5)
```

```
                  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
```

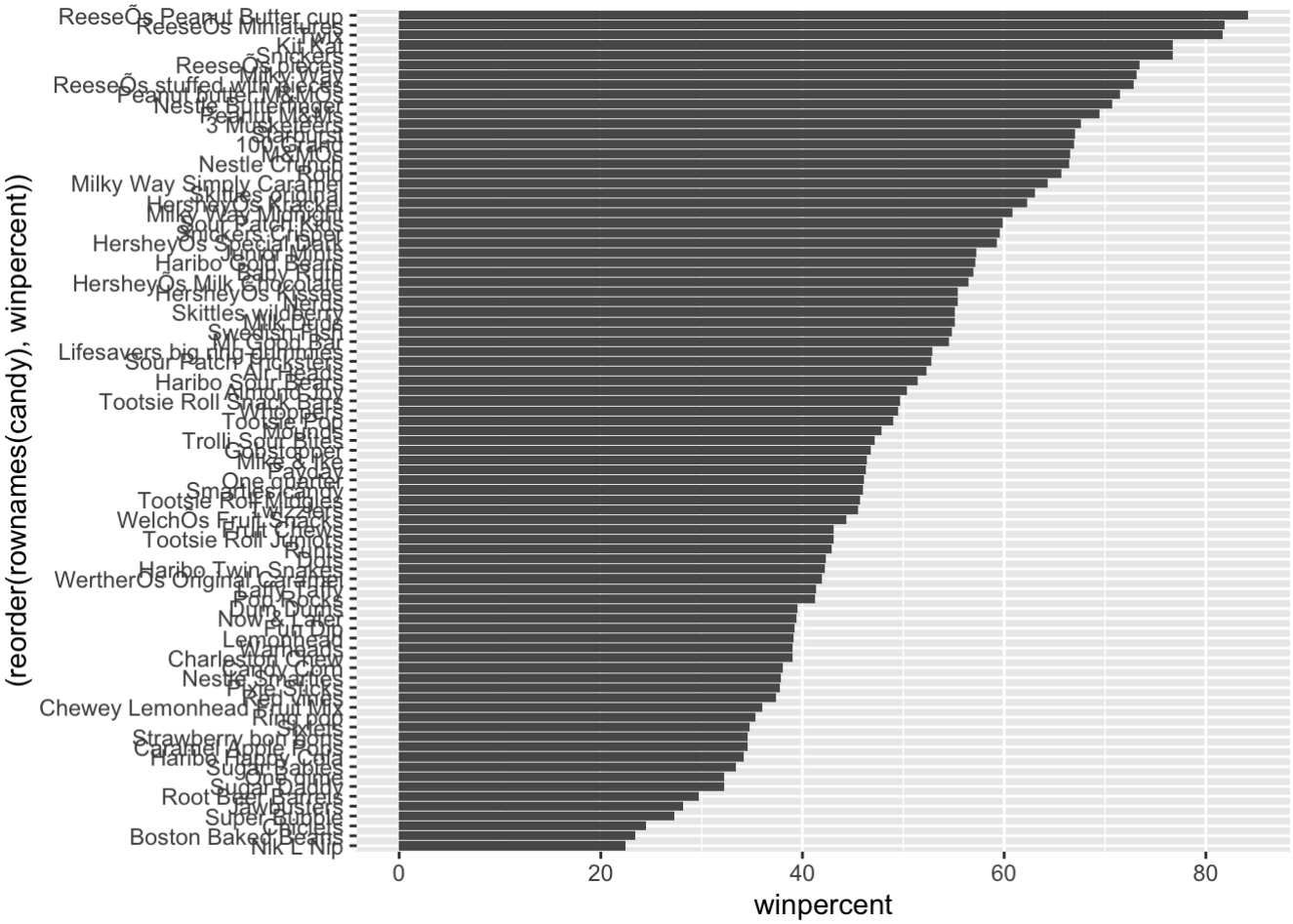
> Q15. Make a first barplot of candy ranking based on winpercent values. Use geom col so all columns are an individual variable and plot

```
ggplot2:: ggplot(candy, aes(x=winpercent, y=rownames(candy))) + geom_col()
```



Lets ReORder these

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

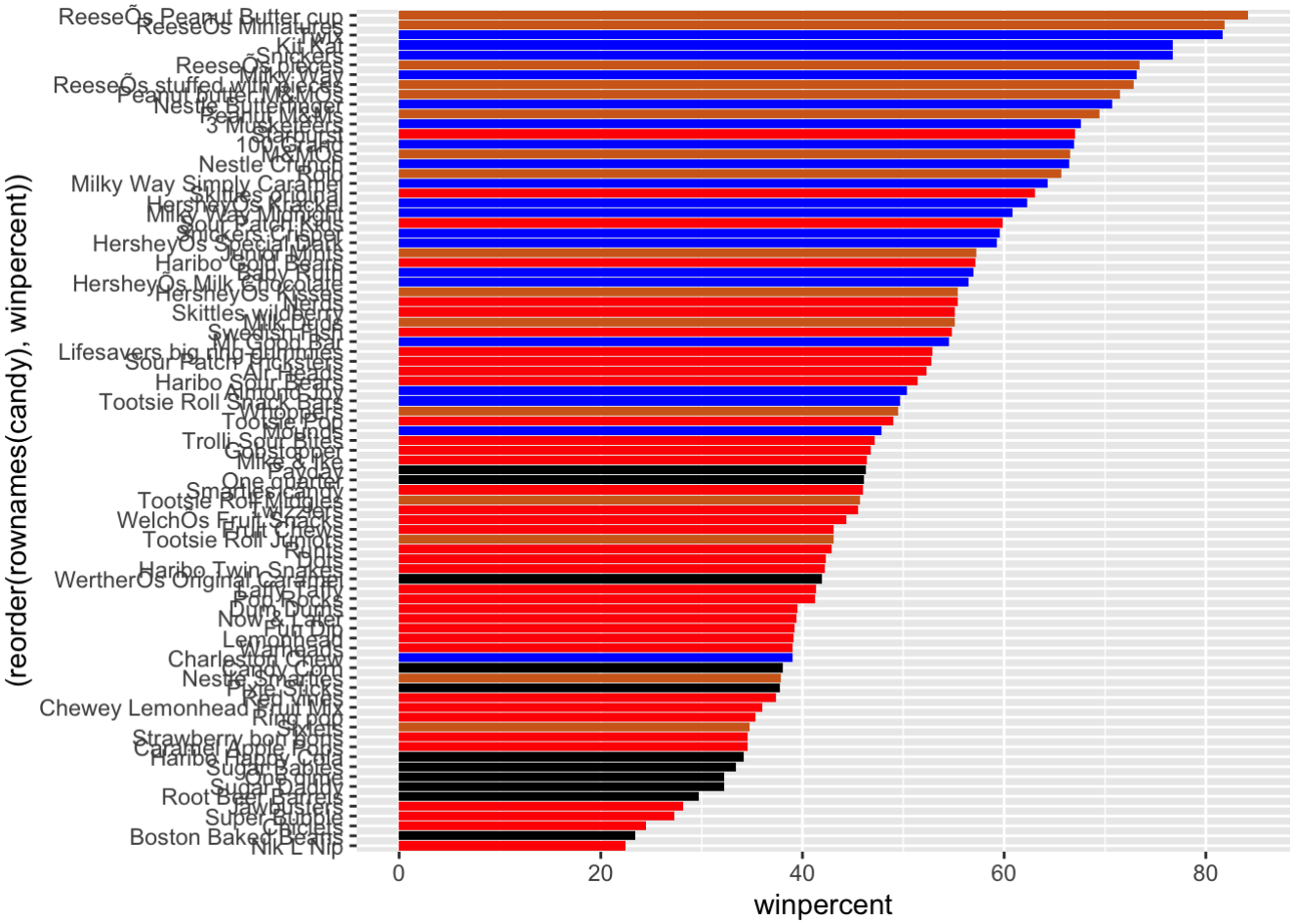


Lets set a color vector (by candy type) that we can Color for every bar by candy type nwor makes a black for every row in candy df

```
my_cols <- rep("black", nrow(candy))
#Makes all rows black
#Then we say, for every TRUE this is a chocolate, then color it chocolate
my_cols[ as.logical(candy$chocolate)] <- "chocolate"
my_cols[ as.logical(candy$bars)] <- "brown"
my_cols[ as.logical(candy$chocolate & candy$bar)] <-"blue"
my_cols[ as.logical(candy$fruity)] <- "red"
```

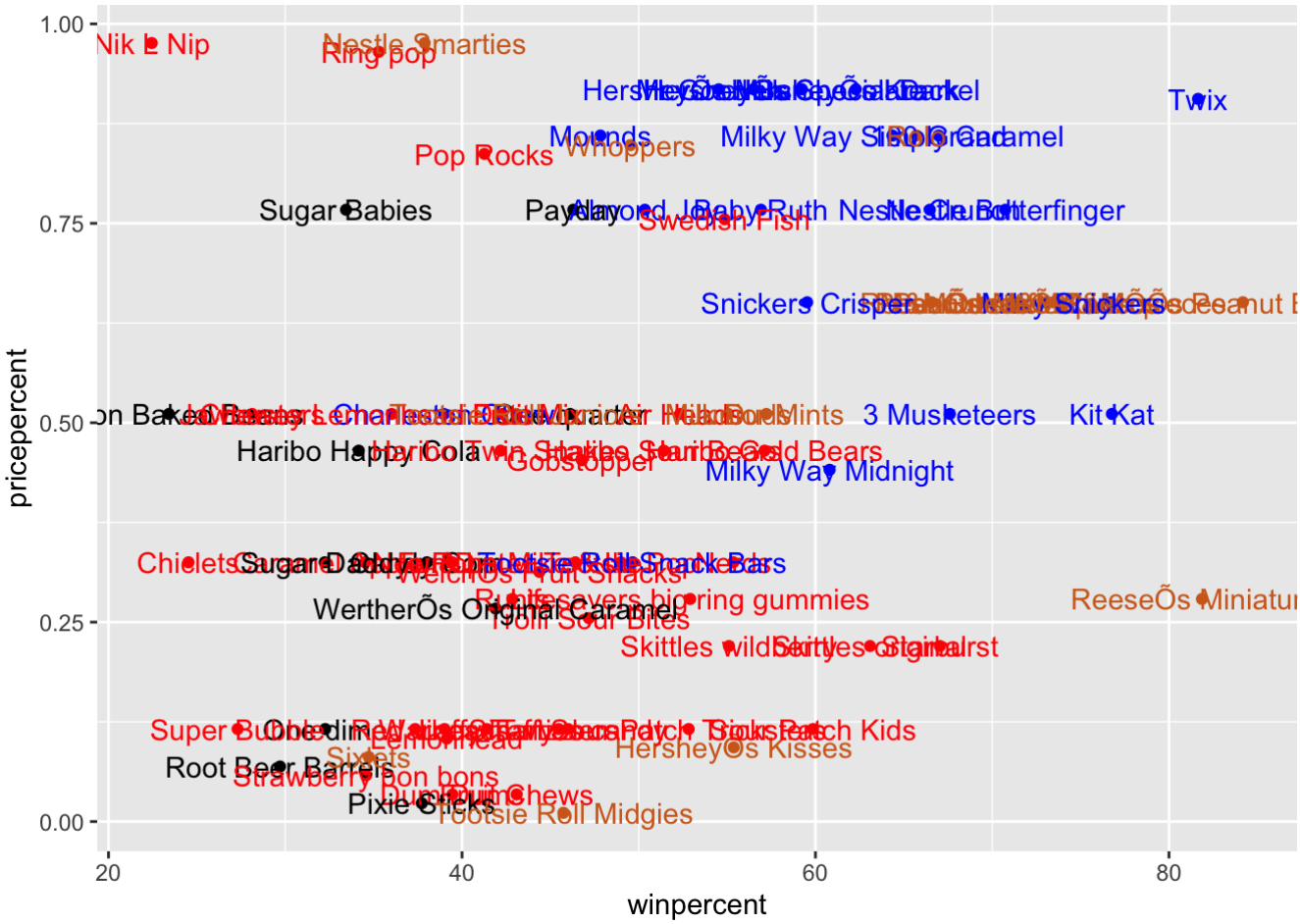
> Q16, order them by winpercent and graph Q17. What is the worst ranked chocolate candy? Sixlets Q18. What is the best ranked fruity candy? Starburst

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



#Looking at pricepercent value for money? what is the best candy for the least money a way to do that is 'winpercent' v. pricepoercent

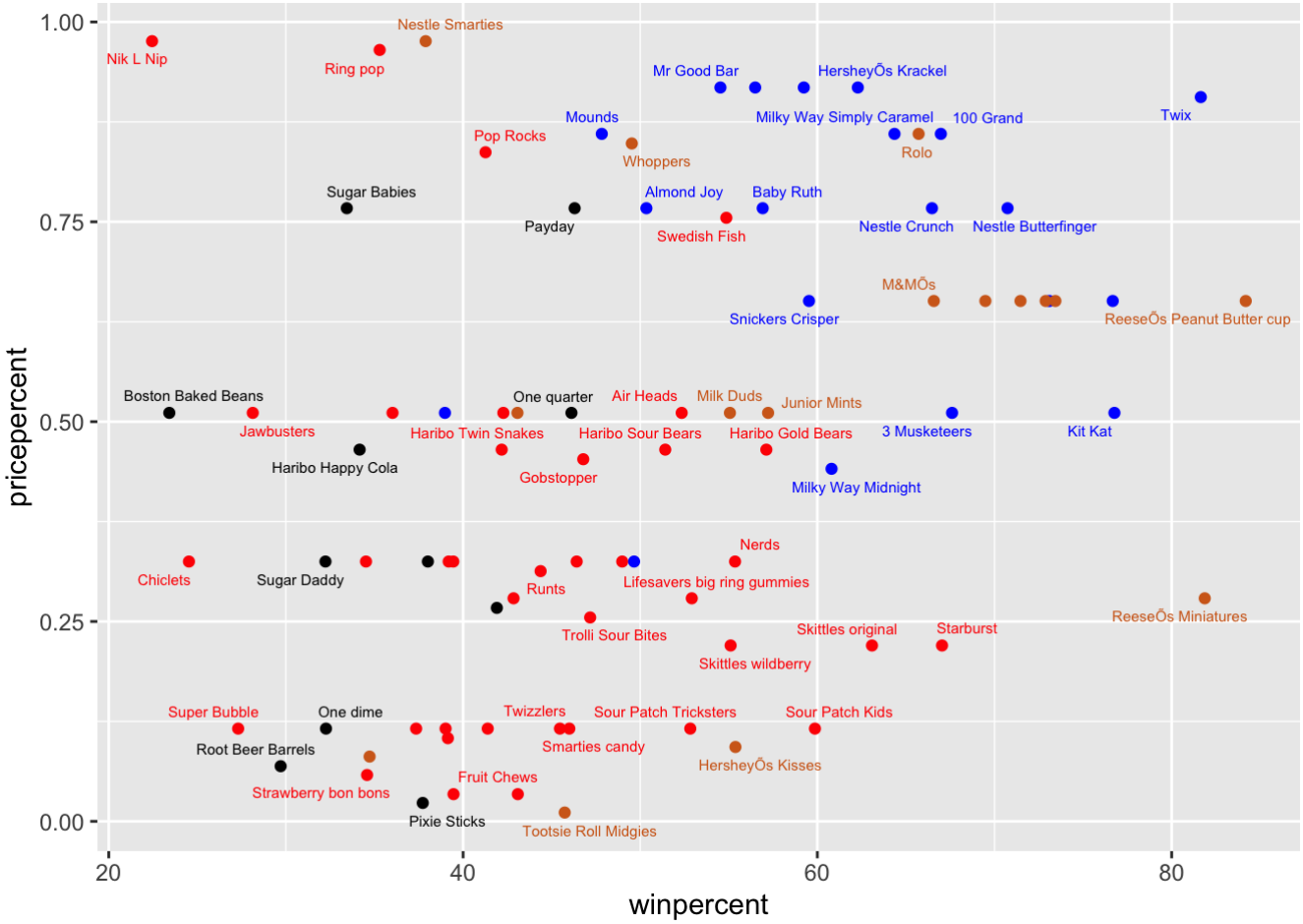
```
ggplot(candy) +
  aes(winpercent, pricepercent, label=rownames(candy)) +
  geom_point(col=my_cols) + geom_text(col=my_cols)
```



Make these points not overlap as much with gg repel

```
library(ggrepel)

ggplot(candy) +
  aes(winpercent, pricepercent, label=rownames(candy)) +
  geom_point(col=my_cols) +  geom_text_repel(col=my_cols, size=2.0, max.overlaps = 5)
```

```
Warning: ggrepel: 27 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? Recess miniatures Q20. What are the top 5 most expensive candy types in the dataset and of these which is the least popular? Nik L Nip, Nestle Smarties, Ring pop, Hersheys Krakel, and Hersheys chocolate. Least popular is Nik L Nip

```
ord_win <- order(candy$winpercent, decreasing = TRUE)
head(candy [ord_win, c(11,12)])
```

```
                         pricepercent winpercent
ReeseÕs Peanut Butter cup        0.651   84.18029
ReeseÕs Miniatures               0.279   81.86626
Twix                             0.906   81.64291
Kit Kat                          0.511   76.76860
Snickers                         0.651   76.67378
ReeseÕs pieces                   0.651   73.43499
```

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

```
                    pricepercent winpercent
Nik L Nip                  0.976   22.44534
Nestle Smarties            0.976   37.88719
```

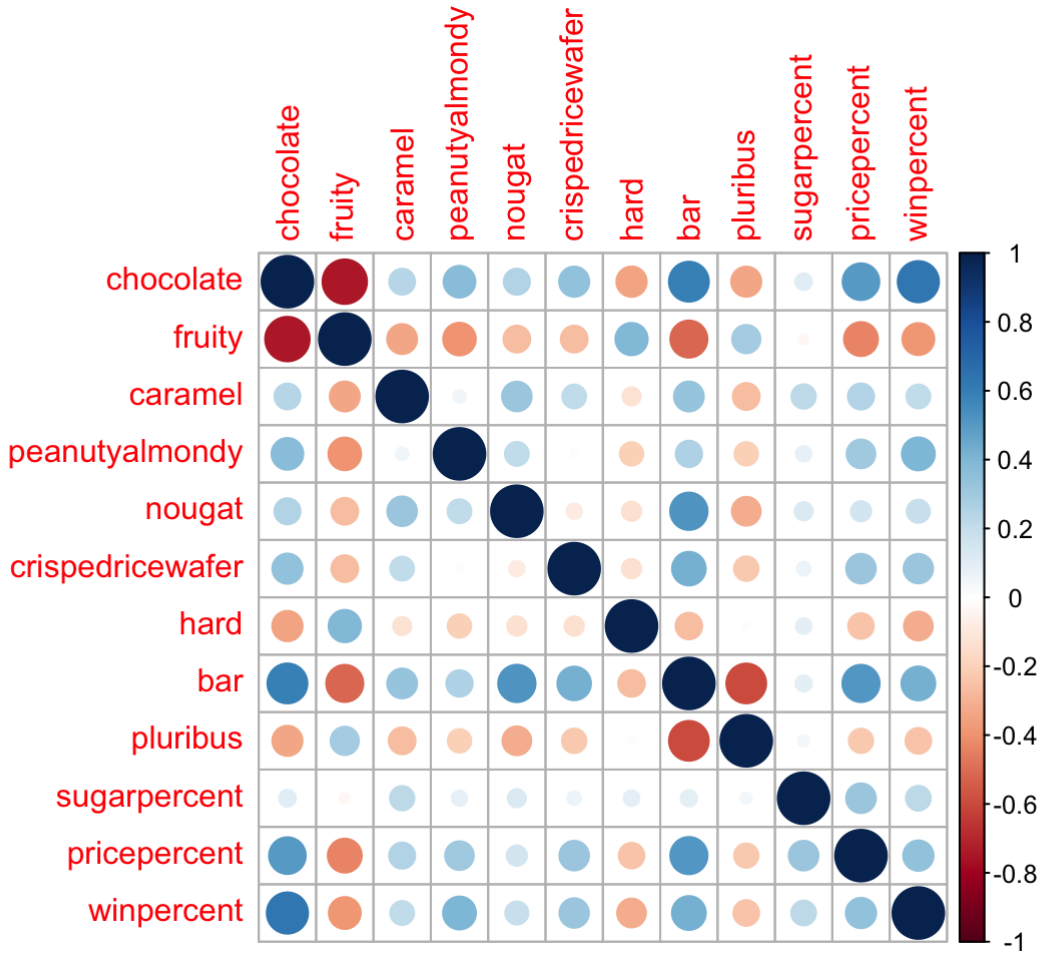
```
Ring pop                      0.965   35.29076
HersheyÕs Krackel             0.918   62.28448
HersheyÕs Milk Chocolate      0.918   56.49050
```

##Sec. 5, exploring coorelation >Q22. Examining this plot what two variables are anti-correlated (i.e. have minus values)? >Q23. Similarly, what two variables are most positively correlated?

```
library(corrplot)
```

```
corrplot 0.92 loaded
```

```
## corrplot 0.90 loaded
cij <- cor(candy)
corrplot(cij)
```



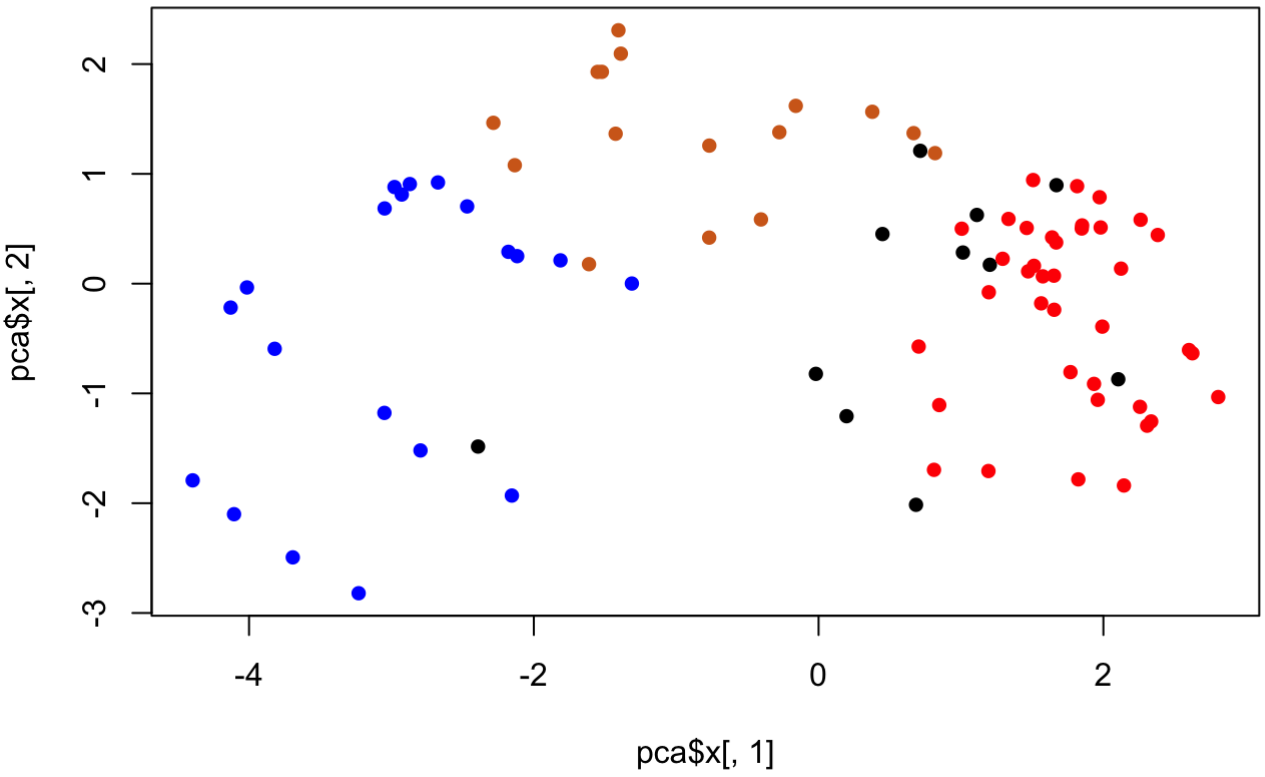
##PCA >Dont forget to set scale to true, its set to scale = false naturally, so that it is scaling varience instead of drastic values, like win percent.

```
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
```

lets plot data on these new axis now. ie pc1 v pc2

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



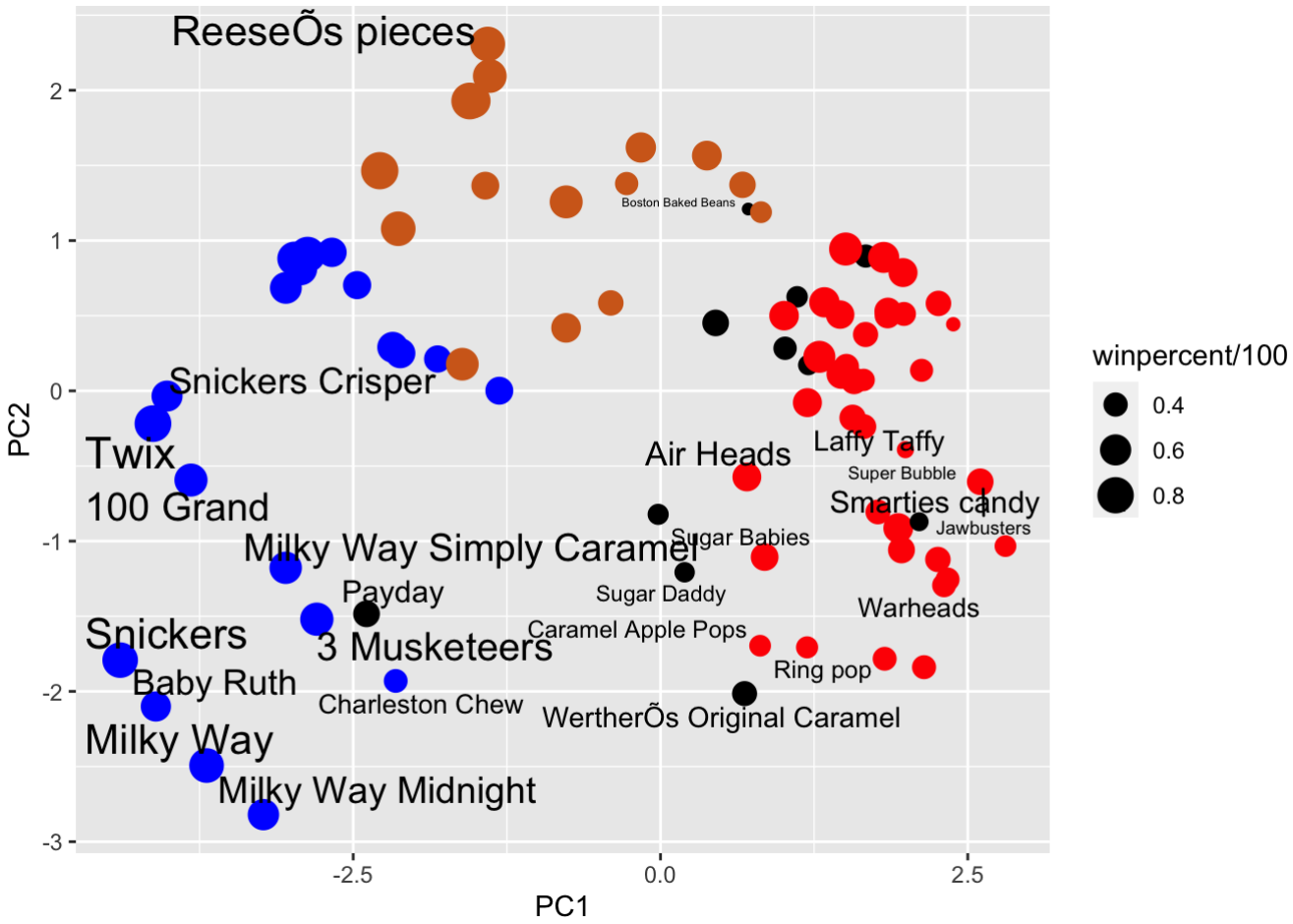
note, gg plot needs a data frame, so need to make a data frame of pca. So we bind the PCA data and add 3 new columns to the cady data frame

```
# Make a new data-frame with our PCA results and candy data
my_data <- cbind(candy, pca$x[,1:3])
```

```
p <- ggplot(my_data) +
        aes(x=PC1, y=PC2,
            size=winpercent/100,
```

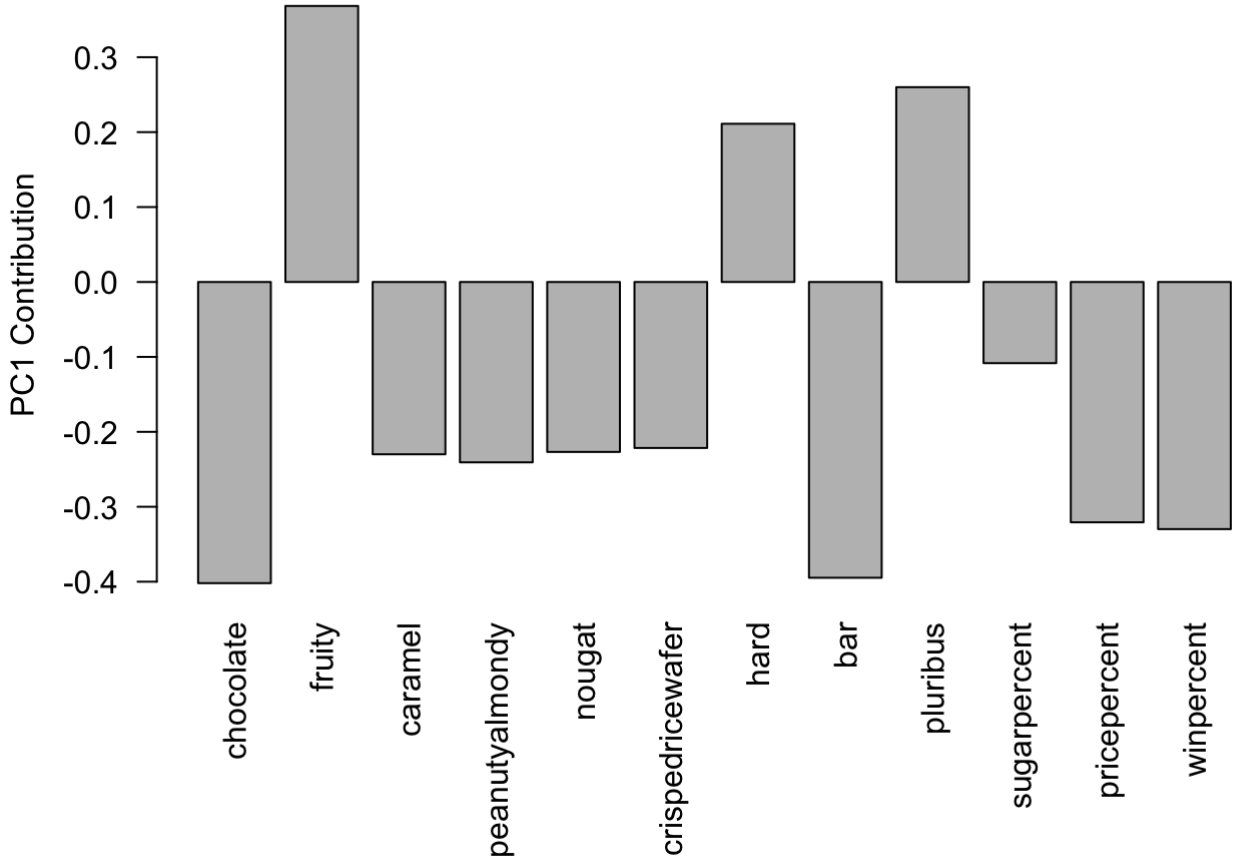
```
            text=rownames(my_data),
            label=rownames(my_data)) +
        geom_point(col=my_cols, ) + geom_text_repel(max.overlaps = 5)

p
```

```
Warning: ggrepel: 61 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```



Can also show how much EACH GROUP, each variable is contributing to the variation making up the PCA axis. This is what the rotation readout is. >Q24.What original variables are picked up strongly by PC1 in the positive direction? Do these make sense to you? Hard, fruity, and comes in multiples seems to pull the variation the most in the positive direction, making up the positive axis of the PC1 axis of the PCA plot.

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


PC1 Contribution
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
chocolate
fruity
caramel
peanutyalmondy
nougat
crispedricewafer
hard
bar
pluribus
sugarpercent
pricepercent
winpercent