

10_28_22_lab

AUTHOR

Andy Tong

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

candy = read.csv(candy_file, 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 0 0	1	0	1
Mounds 0 0	1	0	0
Mr Good Bar 1 0	1	0	0
Nerds 0 0	0	1	0
Nestle Butterfinger 1 0	1	0	0
Nestle Crunch 0 0	1	0	0
Nik L Nip 0 0	0	1	0
Now & Later 0 0	0	1	0
Payday 1 1	0	0	0
Peanut M&Ms 1 0	1	0	0
Pixie Sticks 0 0	0	0	0
Pop Rocks 0 0	0	1	0
Red vines 0 0	0	1	0
Reese's Miniatures 1 0	1	0	0
Reese's Peanut Butter cup 1 0	1	0	0
Reese's pieces 1 0	1	0	0
Reese's stuffed with pieces 1 0	1	0	0
Ring pop 0 0	0	1	0
Rolo 0 0	1	0	1
Root Beer Barrels 0 0	0	0	0
Runts 0 0	0	1	0
Sixlets 0 0	1	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
0.732				0
3 Musketeers		0	0	1
0.604				0
One dime		0	0	0
0.011				0
One quarter		0	0	0
0.011				0
Air Heads		0	0	0
0.906				0
Almond Joy		0	0	1
0.465				0
Baby Ruth		0	0	1
0.604				0
Boston Baked Beans		0	0	0
0.313				1
Candy Corn		0	0	0
0.906				1
Caramel Apple Pops		0	0	0
0.604				0
Charleston Chew		0	0	1
0.604				0
Chewey Lemonhead Fruit Mix		0	0	0
0.732				1
Chiclets		0	0	0
0.046				1
Dots		0	0	0
0.732				1
Dum Dums		0	1	0
0.732				0
Fruit Chews		0	0	0
0.127				1
Fun Dip		0	1	0
0.732				0
Gobstopper		0	1	0
0.906				1
Haribo Gold Bears		0	0	0
0.465				1
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 0.604	0	0	1	0
Nestle Crunch 0.313	1	0	1	0
Nik L Nip 0.197	0	0	0	1
Now & Later 0.220	0	0	0	1
Payday 0.465	0	0	1	0
Peanut M&Ms 0.593	0	0	0	1
Pixie Sticks 0.093	0	0	0	1
Pop Rocks 0.604	0	1	0	1
Red vines 0.581	0	0	0	1
Reese's Miniatures 0.034	0	0	0	0
Reese's Peanut Butter cup 0.720	0	0	0	0
Reese's pieces 0.406	0	0	0	1
Reese's stuffed with pieces 0.988	0	0	0	0
Ring pop 0.732	0	1	0	0
Rolo 0.860	0	0	0	1
Root Beer Barrels 0.732	0	1	0	1
Runts 0.872	0	1	0	1
Sixlets 0.220	0	0	0	1
Skittles original 0.941	0	0	0	1
Skittles wildberry 0.941	0	0	0	1
Nestle Smarties 0.267	0	0	0	1
Smarties candy 0.267	0	1	0	1
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&M's	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?

```
cat("There are", nrow(candy), "types of candy in this dataset")
```

There are 85 types of candy in this dataset

Q2. How many fruity candy types are in the dataset? The functions `dim()`, `nrow()`, `table()` and `sum()` may be useful for answering the first 2 questions.

```
cat("There are", sum(candy$fruity), "fruity candies")
```

There are 38 fruity candies

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

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

```
[1] 63.08514
```

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

Data summary

Name	candy
Number of rows	85
Number of columns	12
<hr/>	
Column type frequency:	
numeric	12

Group variables	None
-----------------	------

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25
chocolate	0	1	0.44	0.50	0.00	0.00
fruity	0	1	0.45	0.50	0.00	0.00
caramel	0	1	0.16	0.37	0.00	0.00
peanutyalmondy	0	1	0.16	0.37	0.00	0.00
nougat	0	1	0.08	0.28	0.00	0.00
crispedricewafer	0	1	0.08	0.28	0.00	0.00
hard	0	1	0.18	0.38	0.00	0.00
bar	0	1	0.25	0.43	0.00	0.00
pluribus	0	1	0.52	0.50	0.00	0.00
sugarpercent	0	1	0.48	0.28	0.01	0.22
pricepercent	0	1	0.47	0.29	0.01	0.26
winpercent	0	1	50.32	14.71	22.45	39.14

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 is on a very different scale. It goes from 0 -> 100 while other columns go from 0 -> 1

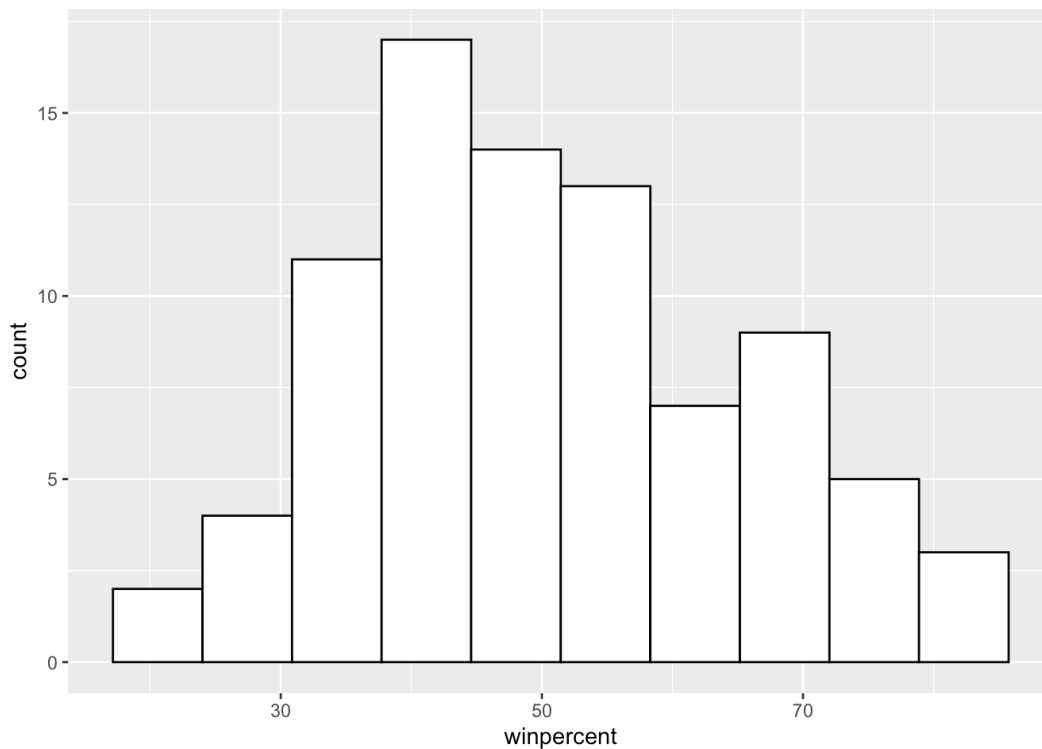
Q7. What do you think a zero and one represent for the candy\$chocolate column?

A 0 means that this candy does not have chocolate in it. A 1 means that this candy does have chocolate in it.

Q8. Plot a histogram of winpercent values

```
library(ggplot2)
p<-ggplot(candy, aes(x=winpercent)) +
```

```
geom_histogram(bins = 10, color="black", fill="white")
p
```



Q9. Is the distribution of winpercent values symmetrical?

Not completely. The left of the peak seems to drop off a lot more than the right side which drops at a much lower rate. Also the peak lies below 50%.

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

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

```
mean_choco_win <- mean(candy[candy$chocolate == 1,]$winpercent)
mean_fruit_win <- mean(candy[candy$fruity == 1,]$winpercent)
cat("The mean chocolate win percent is", mean_choco_win, "while
```

The mean chocolate win percent is 60.92153 while the fruity mean win percent is 44.11974

```
cat("\nThis tells us that to say chocolates are on average ranked
```

This tells us that to say chocolates are on average ranked

higher than fruity candy is TRUE

Q12. Is this difference statistically significant?

```
t.test(candy[candy$chocolate == 1,]$winpercent, candy[candy$fru
```

Welch Two Sample t-test

```
data: candy[candy$chocolate == 1,]$winpercent and
candy[candy$fruity == 1,]$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
```

Yes it's pretty significant with a p value of 2.871×10^{-8}

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

```
ordered_candy <- candy[order(candy$winpercent),]
lowest_5 <- rownames(ordered_candy)[1:5]
print("The 5 least liked candies were ")
```

```
[1] "The 5 least liked candies were "
```

```
print(lowest_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?

```
highest_5 <- rownames(ordered_candy)[(nrow(ordered_candy)-4):n
print("The 5 most liked candies were ")
```

```
[1] "The 5 most liked candies were "
```

```
for(candy_name in highest_5){
  print(candy_name)
}
```

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

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

```
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
0.732				0
3 Musketeers		0	0	1
0.604				0
One dime		0	0	0
0.011				0
One quarter		0	0	0
0.011				0
Air Heads		0	0	0
0.906				0
Almond Joy		0	0	1
0.465				0
Baby Ruth		0	0	1
0.604				0
Boston Baked Beans		0	0	0
0.313				1
Candy Corn		0	0	0
0.906				1
Caramel Apple Pops		0	0	0
0.604				0
Charleston Chew		0	0	1
0.604				0
Chewey Lemonhead Fruit Mix		0	0	0
0.732				1
Chiclets		0	0	0
0.046				1
Dots		0	0	0
0.732				1
Dum Dums		0	1	0
0.732				0
Fruit Chews		0	0	0
0.127				1
Fun Dip		0	1	0
0.732				0
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

```
p1 <- ggplot(candy)+
  aes(winpercent, reorder(rownames(candy),winpercent), height =
```



```
geom_bar(stat="identity", width=1.5)
```

```
ggsave("candy_winpercent_barplot.png", plot = p1, height = 10)
```

Saving 7 x 10 in image

Warning: position_stack requires non-overlapping x intervals

```
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"
```

```
p2 <- ggplot(candy) +
  aes(winpercent, reorder(rownames(candy), winpercent)) +
  geom_col(fill=my_cols)
```

```
ggsave("candy_winpercent_barplot_colored.png", plot = p2, height = 12)
```

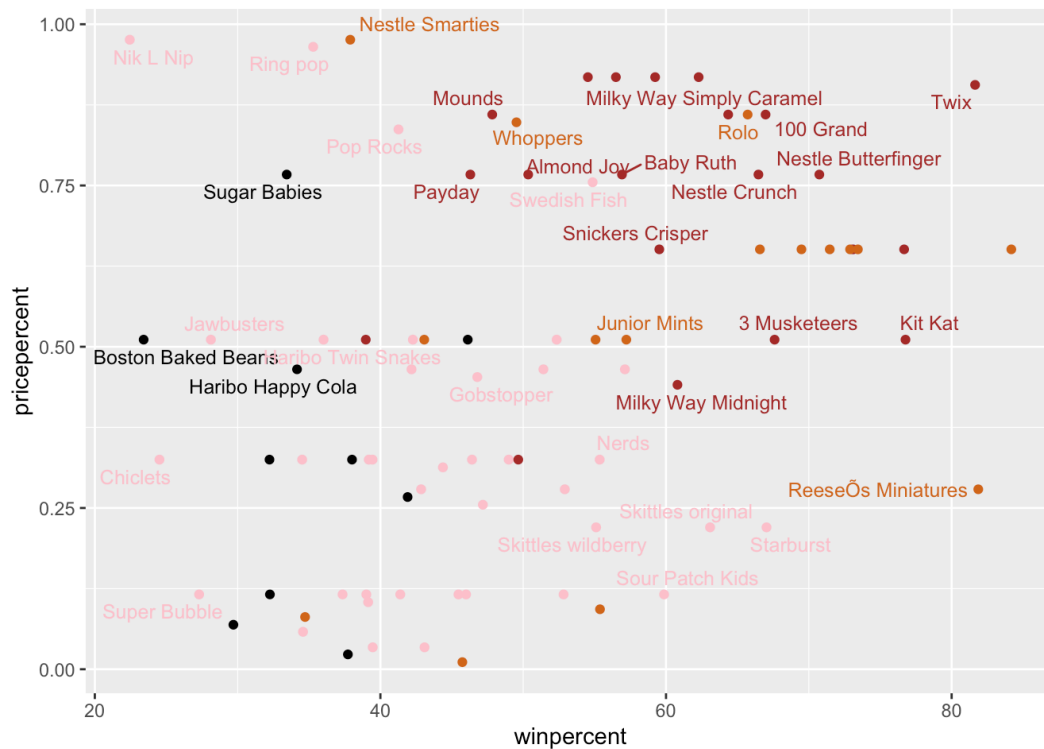
Saving 7 x 12 in image

- Q17. What is the worst ranked chocolate candy? Sixlets are the worst ranked chocolate candy
- Q18. What is the best ranked fruity candy? Starbursts are the best ranked fruity candy

```
library(ggrepel)

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

Warning: ggrepel: 50 unlabeled data points (too many overlaps). Consider increasing max.overlaps



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

This has an arguable answer depending on what one's optimization function is of winpercent and money. The answer could probably be one of toostie roll midgies, ReeseOs Miniatures, ReeseOs Peanut Butter cup, and maybe sour patch kids. I believe each of these could claim a line that represents an optimization function of a winpercent to price tradeoff function. ReeseOs Miniatures is probably the most realistic answer, but someone that only cares about winpercent could argue ReeseOs Peanut Butter cup while another that optimizes extremely for price could go for toostie roll midgies which have the lowest price.

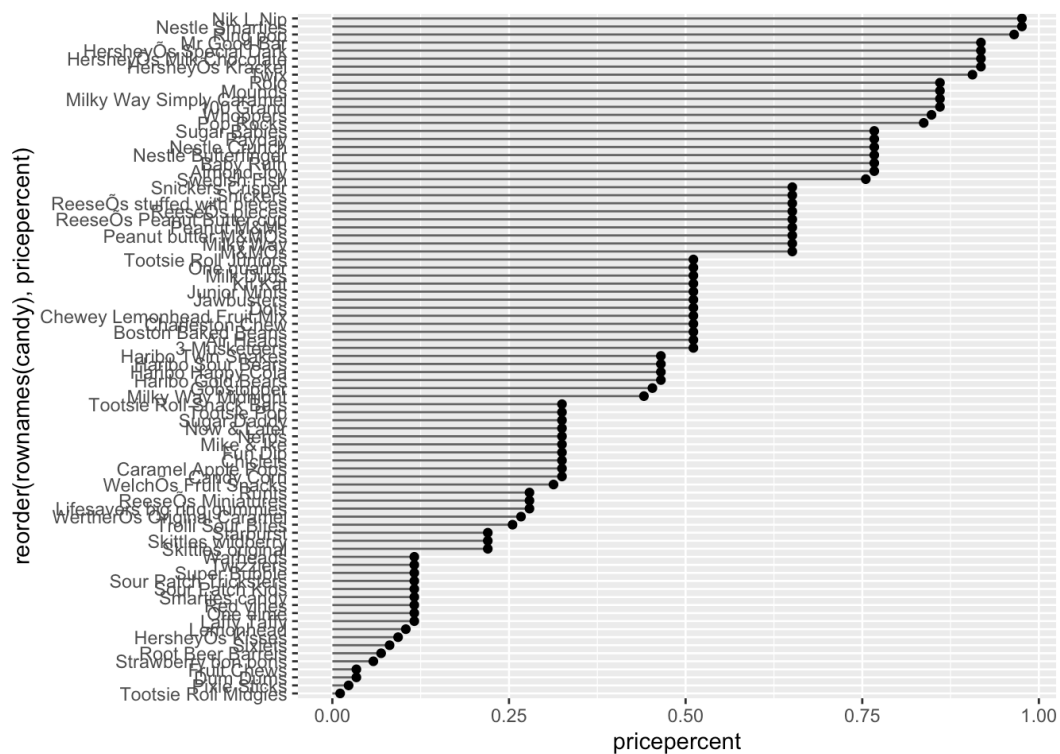
Q20. What are the top 5 most expensive candy types in the dataset and of these which is the least popular? The top 5 most expensive are Nik L Nip, Nestle Smarties, Ring Pops, Mr Good Bar, and HersheyOs Special Dark. Among these, the least popular is Nik L Nip

Q21. Make a barplot again with `geom_col()` this time using pricepercent and then improve this step by step, first ordering the x-axis by value and finally making a so called "dot chat" or "lollipop" chart by swapping `geom_col()` for `geom_point()` + `geom_segment()`.

```
p3 <- ggplot(candy) +
  aes(pricepercent, reorder(rownames(candy), pricepercent)) +
  geom_col(fill=my_cols)
ggsave("candy_price_percent_barplot_colored.png", plot = p3, height = 12, width = 7)
```

Saving 7 x 12 in image

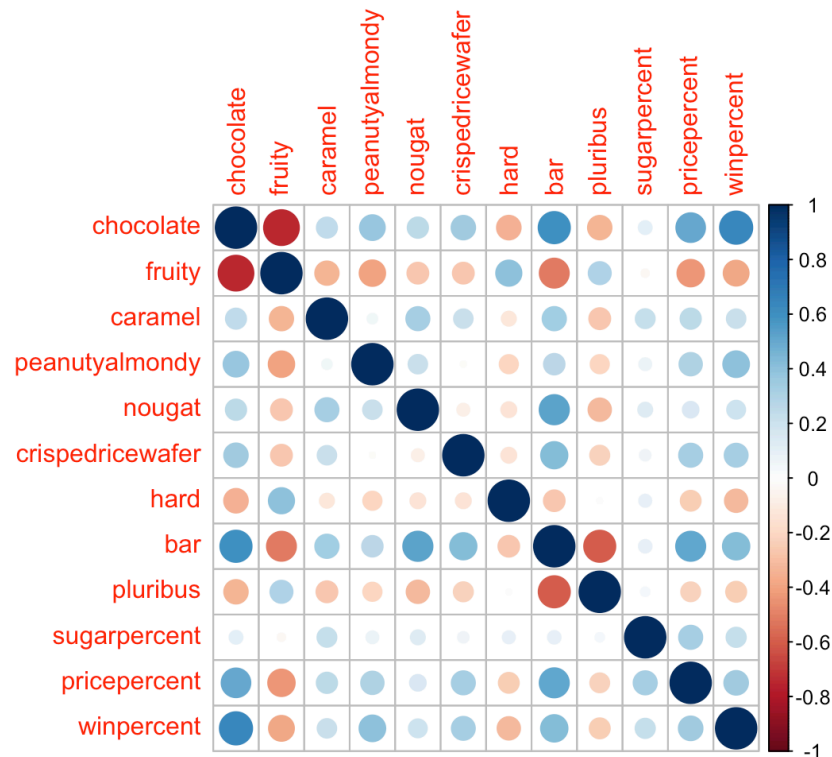
```
ggplot(candy) +
  aes(pricepercent, reorder(rownames(candy), pricepercent)) +
  geom_segment(aes(yend = reorder(rownames(candy), pricepercent),
    xend = 0), col="gray40") +
  geom_point()
```



```
library(corrplot)
```

corrplot 0.92 loaded

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



Q22. Examining this plot what two variables are anti-correlated (i.e. have minus values)? The three with the biggest anti correlation in order from most anti correlated to least are fruity and chocolate, then pluribus and bar, and then bar and fruity. So chocolate and fruit snacks don't often get put together, then bars usually come in packages of 1, and then bars often are not made of fruity materials.

Q23. Similarly, what two variables are most positively correlated? The two that seem the most correlated are chocolate and winpercent and chocolate and bars. That means that chocolates are very popular and they often come in bar form.

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

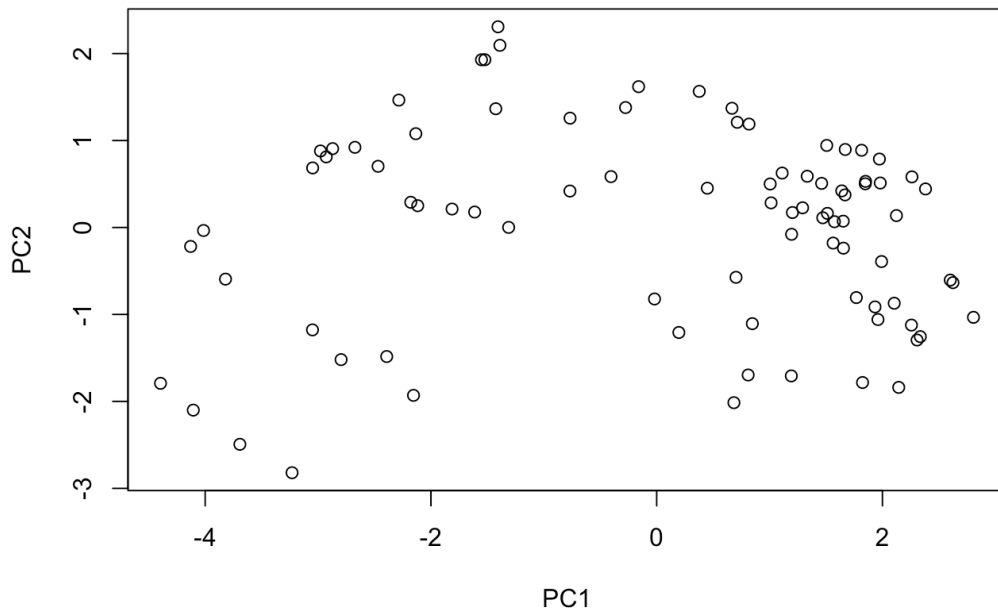
Importance of components:

	PC1	PC2	PC3	PC4	PC5
PC6					
PC7					
Standard deviation	2.0788	1.1378	1.1092	1.07533	0.9518
	0.81923	0.81530			
Proportion of Variance	0.3601	0.1079	0.1025	0.09636	0.0755
	0.05593	0.05539			
Cumulative Proportion	0.3601	0.4680	0.5705	0.66688	0.7424

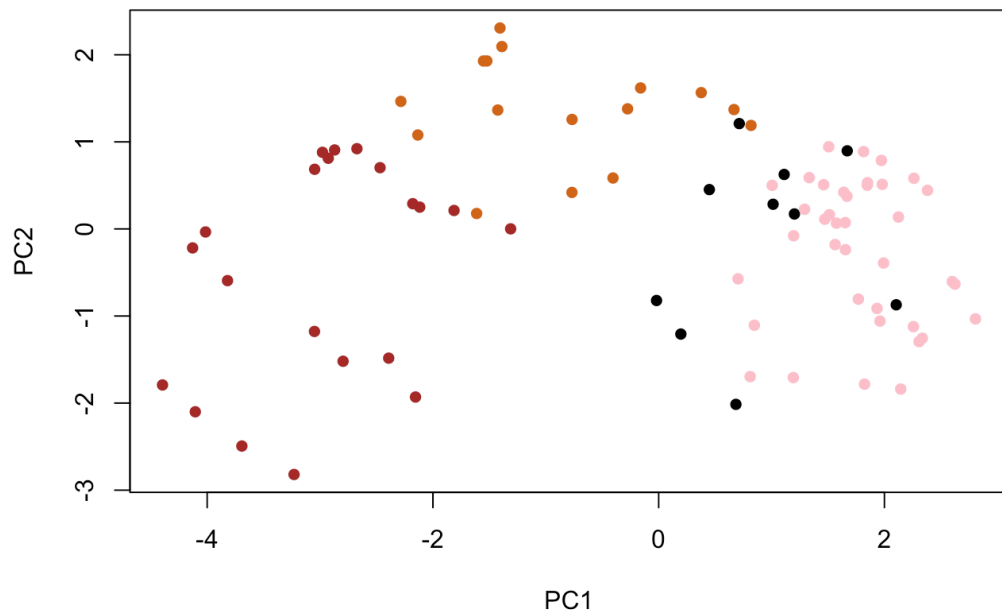
0.79830 0.85369

	PC8	PC9	PC10	PC11	PC12
Standard deviation	0.74530	0.67824	0.62349	0.43974	0.39760
Proportion of Variance	0.04629	0.03833	0.03239	0.01611	0.01317
Cumulative Proportion	0.89998	0.93832	0.97071	0.98683	1.00000

```
plot(pca$x[,1], pca$x[,2], xlab="PC1", ylab="PC2")
```

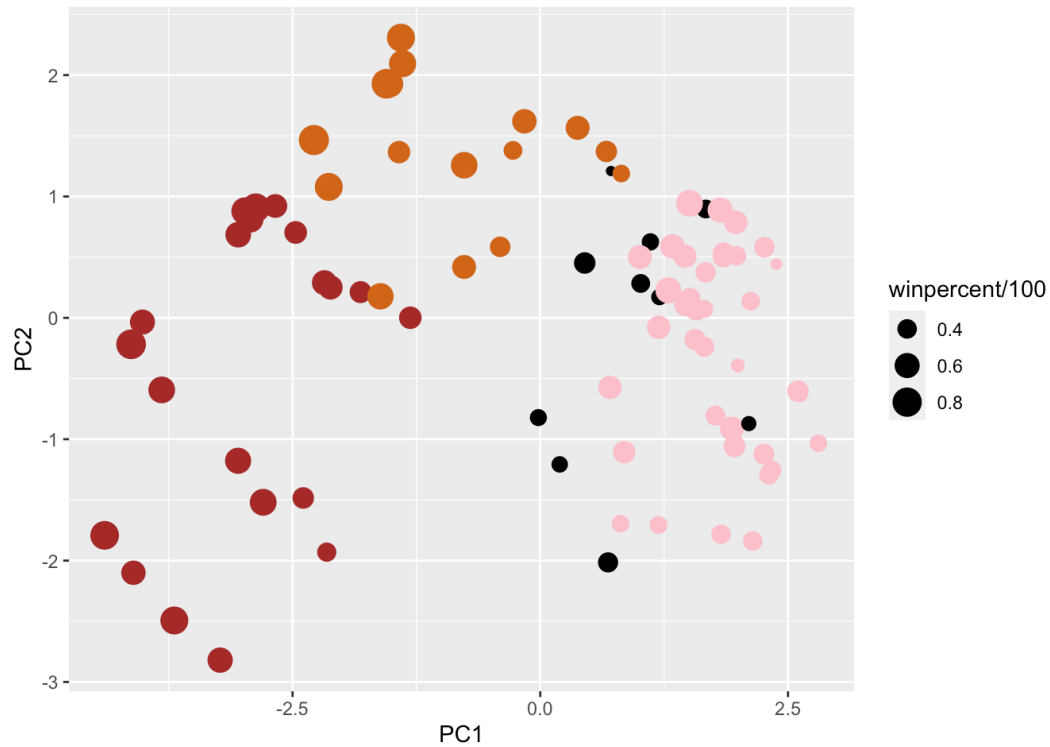


```
plot(pca$x[,1:2], col=my_cols, pch=16)
```



```
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)
```

p



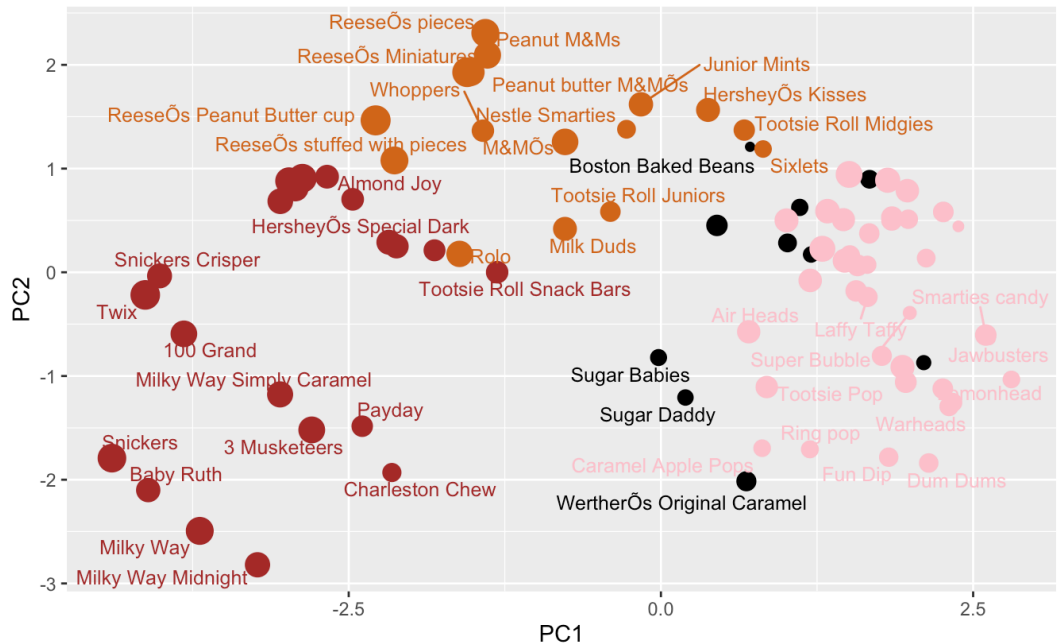
```
library(ggrepel)
```

```
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), c",  
        caption="Data from 538")
```

Warning: ggrepel: 39 unlabeled data points (too many overlaps). Consider increasing max.overlaps

Halloween Candy PCA Space

Colored by type: chocolate bar (dark brown), chocolate other (light brown), fruity (red), other (black)



Data from 538

```
library(plotly)
```

Attaching package: 'plotly'

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

```
last_plot
```

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

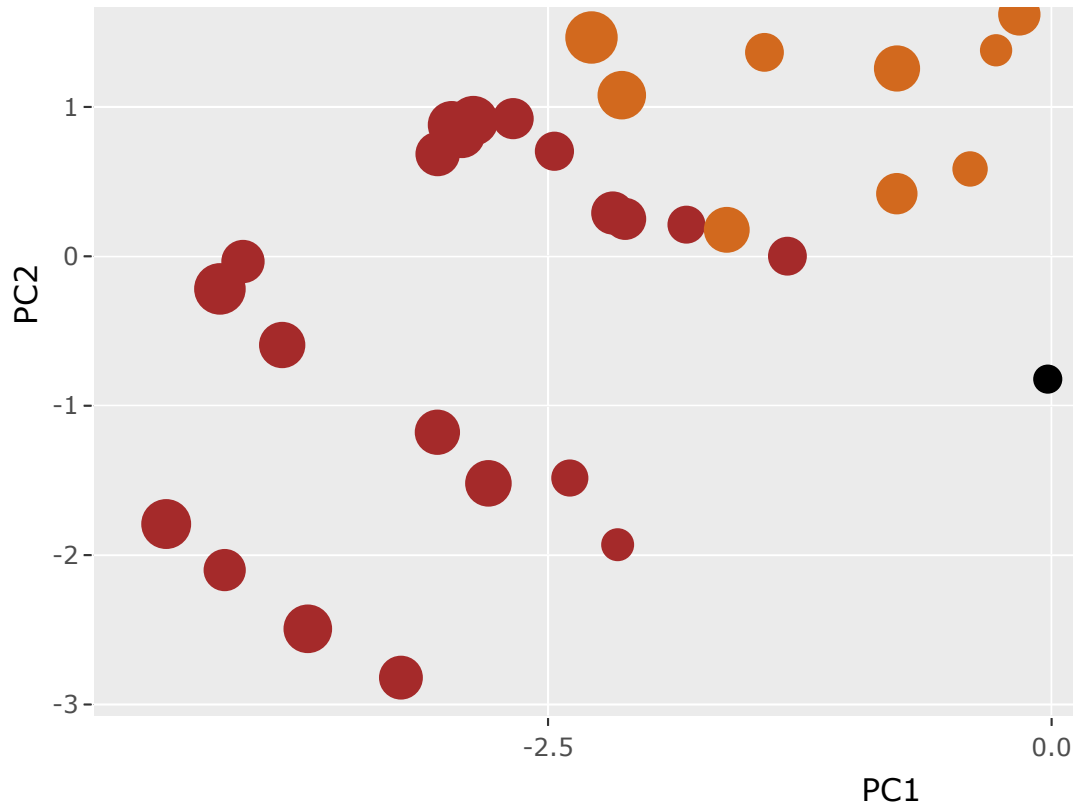
```
filter
```

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

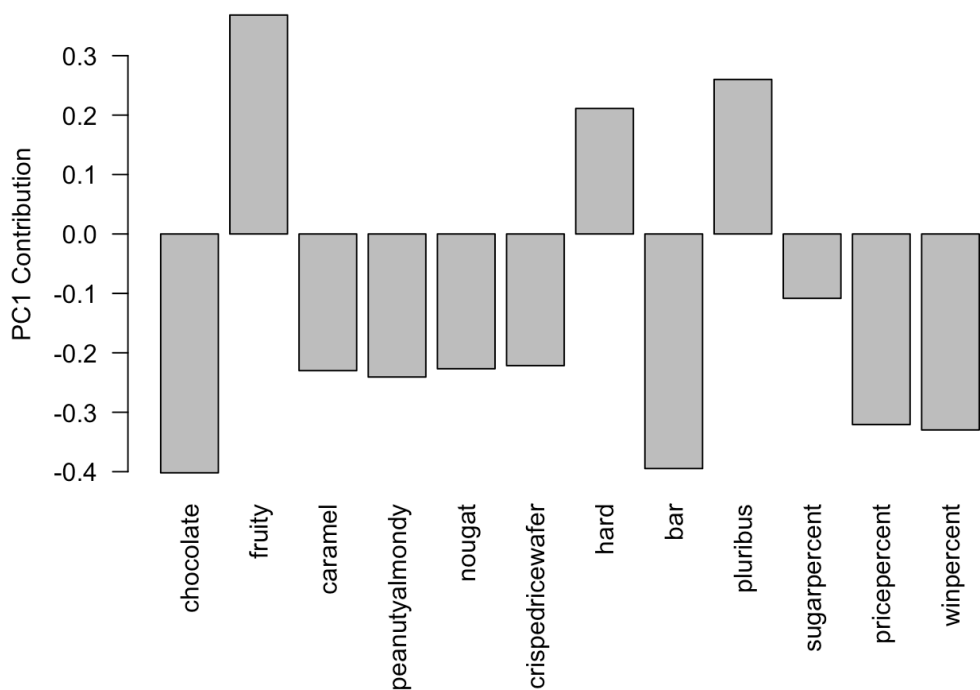
```
layout
```

```
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?

The 3 that have the biggest impact on PC1 are chocolate, fruity, and bar. Yeah these make a lot of sense that the biggest contributing factor to separating out 3 groups are by the overall largest category that describes the type of candy they will be. A candy's overall type will decide a number of things about the candy's attributes.