

Class 09 Halloween

Abraham Rachlin

Table of contents

Exploratory Analysis	3
3. Overall Candy Rankings.	6
4. Taking a look at pricepercent	17

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

```
candy = read.csv(candy_file, row.names=1)
head(candy)
```

	chocolate	fruity	caramel	peanutyalmondy	nougat	crispedricewafer
100 Grand	1	0	1	0	0	1
3 Musketeers	1	0	0	0	1	0
One dime	0	0	0	0	0	0
One quarter	0	0	0	0	0	0
Air Heads	0	1	0	0	0	0
Almond Joy	1	0	0	1	0	0

	hard	bar	pluribus	sugarpercent	pricepercent	winpercent
100 Grand	0	1	0	0.732	0.860	66.97173
3 Musketeers	0	1	0	0.604	0.511	67.60294
One dime	0	0	0	0.011	0.116	32.26109
One quarter	0	0	0	0.011	0.511	46.11650
Air Heads	0	0	0	0.906	0.511	52.34146
Almond Joy	0	1	0	0.465	0.767	50.34755

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

```
nrow(candy)
```

```
[1] 85
```

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

```
candy$fruity
```

```
[1] 0 0 0 0 1 0 0 0 0 1 0 1 1 1 1 1 1 1 0 1 1 0 0 0 0 1 0 0 1 1 1 0 0 1 0 0 0  
[39] 0 0 0 1 0 0 1 1 0 0 0 1 1 0 0 0 0 1 0 0 1 0 1 1 0 1 0 0 1 1 1 1 0 0 1 1 1 0  
[77] 0 0 1 0 1 1 1 0 0
```

```
table(candy$fruity)
```

```
 0  1  
47 38
```

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

```
[1] 81.64291
```

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

My favorite candy is Twix.

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

```
[1] 81.64291
```

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
```

Exploratory Analysis

We can use the **skimr** package to get a quick overview of a given dataset. This can be useful for the first time you encounter a new dataset. Now, we will input the skimr package.

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

Table 1: Data summary

Name	candy
Number of rows	85
Number of columns	12
<hr/>	
Column type frequency: numeric	12
<hr/>	
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 variable that appears to be on a different scale to the majority, I would say, is the **winpercent** variable, since it appears to be whole numbers versus all the other variables being decimal.

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

```
candy$chocolate
```

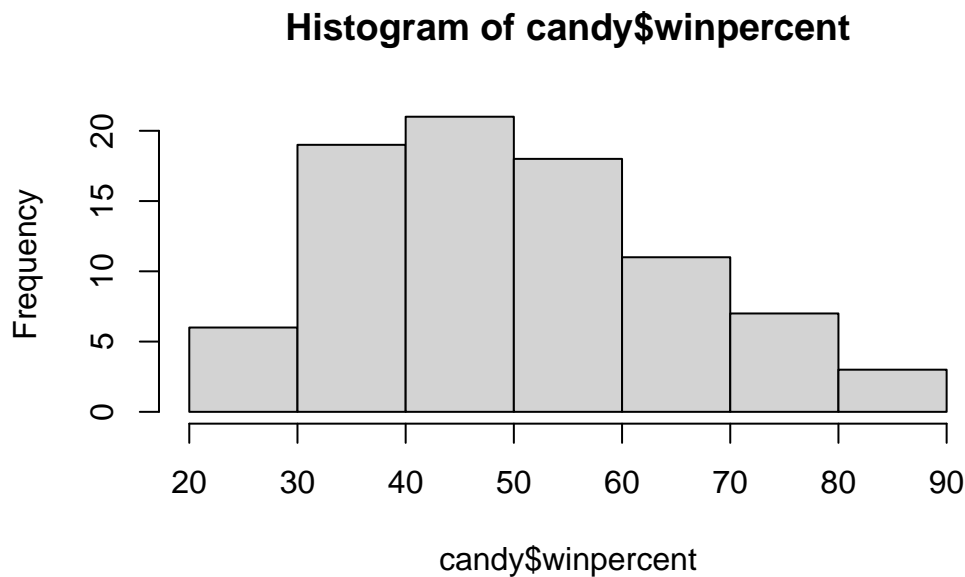
```
[1] 1 1 0 0 0 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 1 1 1 1 0 1 1 0 0 0 1 1 0 1 1 1  
[39] 1 1 1 0 1 1 0 0 0 1 0 0 0 1 1 1 1 0 1 0 0 1 0 0 1 0 1 1 0 0 0 0 0 0 0 1 1  
[77] 1 1 0 1 0 0 0 0 1
```

I believe the one and zero would represent if the candy has chocolate in it, as not all candy types are chocolate-based.

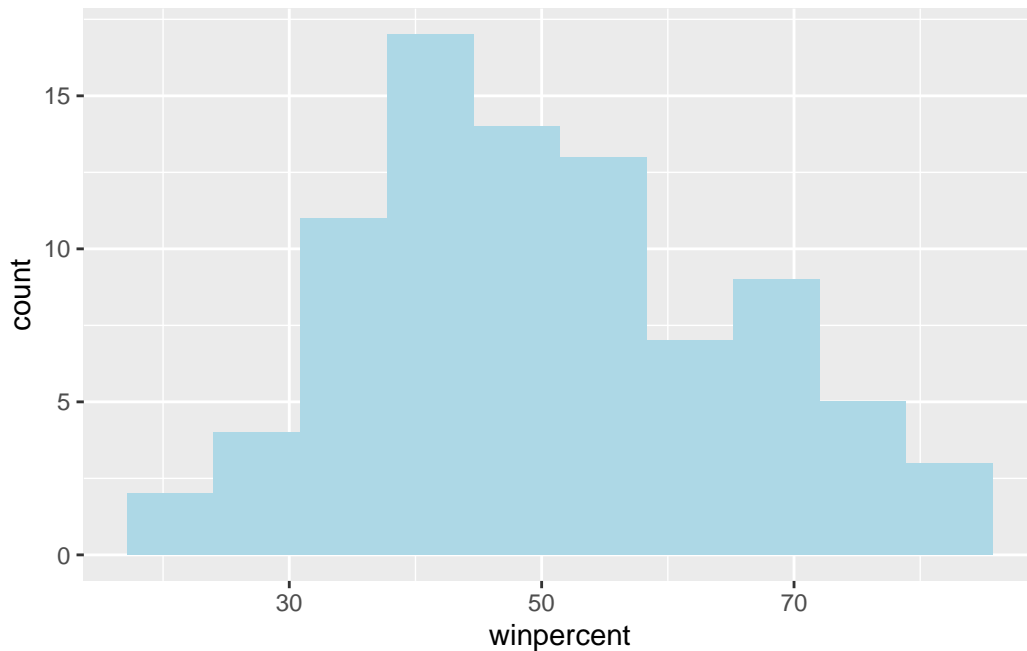
Now, a ggplot of a histogram will be used to visualize the data for explanatory analysis.

Q8. Plot a histogram of winpercent values.

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



```
library(ggplot2)  
ggplot(candy) +  
  aes(winpercent) +  
  geom_histogram(bins = 10, fill="lightblue")
```



Q9. Is the distribution of winpercent values symmetrical?

No, the distribution is not symmetrical.

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

Median is lower than 50%, but the mean is higher than 50%, so, if following the median, is no.

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

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

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

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

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

The mean for the fruity candy is much lower than that of chocolate, as the chocolate is about 60.92, while fruity is 44.12.

Q12. Is this difference statistically significant?

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

Welch Two Sample t-test

```
data:  choc.win and fruit.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
```

It has a p-value below 0.05, meaning that the results are statistically significant.

3. Overall Candy Rankings.

```
head(candy$winpercent)
```

```
[1] 66.97173 67.60294 32.26109 46.11650 52.34146 50.34755
```

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

There are two related functions that can help here, one is the classic `sort()` and `order()`

```
x <- c(5,10,1,4)
sort(x, decreasing = T)
```

```
[1] 10  5  4  1
```

```
order(x)
```

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

```
inds <- order(candy$winpercent)
candy[inds,]
```

	chocolate	fruity	caramel	peanutyalmondy	nougat
Nik L Nip	0	1	0	0	0
Boston Baked Beans	0	0	0	1	0
Chiclets	0	1	0	0	0
Super Bubble	0	1	0	0	0
Jawbusters	0	1	0	0	0
Root Beer Barrels	0	0	0	0	0
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

Werther's 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
Welch's 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
Hershey's Kisses	1	0	0	0	0
Hershey's 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
Hershey's 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
Hershey's 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&M's	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&M's	1	0	0	1	0	
Reese's stuffed with pieces	1	0	0	1	0	
Milky Way	1	0	1	0	1	
Reese's 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	
Reese's Miniatures	1	0	0	1	0	
Reese's Peanut Butter cup	1	0	0	1	0	
	crisped	ricewafer	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

Werther's 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
Welch's 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
Hershey's Kisses	0	0	0	1	0.127
Hershey's 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
Hershey's 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
Hershey's 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&M's	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&M's	0	0	0	1	0.825
Reese's stuffed with pieces	0	0	0	0	0.988
Milky Way	0	0	1	0	0.604
Reese's 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
Reese's Miniatures	0	0	0	0	0.034
Reese's 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

Werther's 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
Welch's 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
Hershey's Kisses	0.093	55.37545
Hershey's 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
Hershey's 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
Hershey's 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&M's	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&M's	0.651	71.46505
Reese's stuffed with pieces	0.651	72.88790
Milky Way	0.651	73.09956
Reese's pieces	0.651	73.43499
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

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

	chocolate	fruity	caramel	peanut	almond	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

	crisped	rice	wafer	hard	bar	pluribus	sugar	percent	price	percent
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

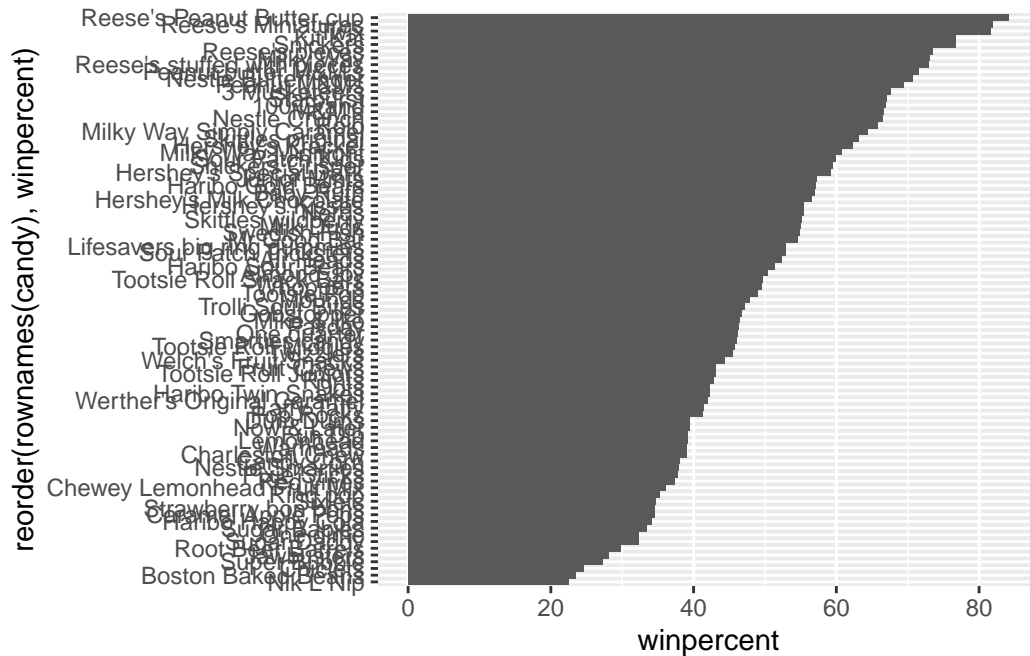
Nik L Nip, Boston Baked Beans, Chiclets, Super Bubble, and Jawbusters.

Q15. What are the top 5 all time favorite candy types out of this set?

Reese's Peanut Butter cup
 Reese's Miniatures
 Twix
 Kit Kat
 Snickers

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

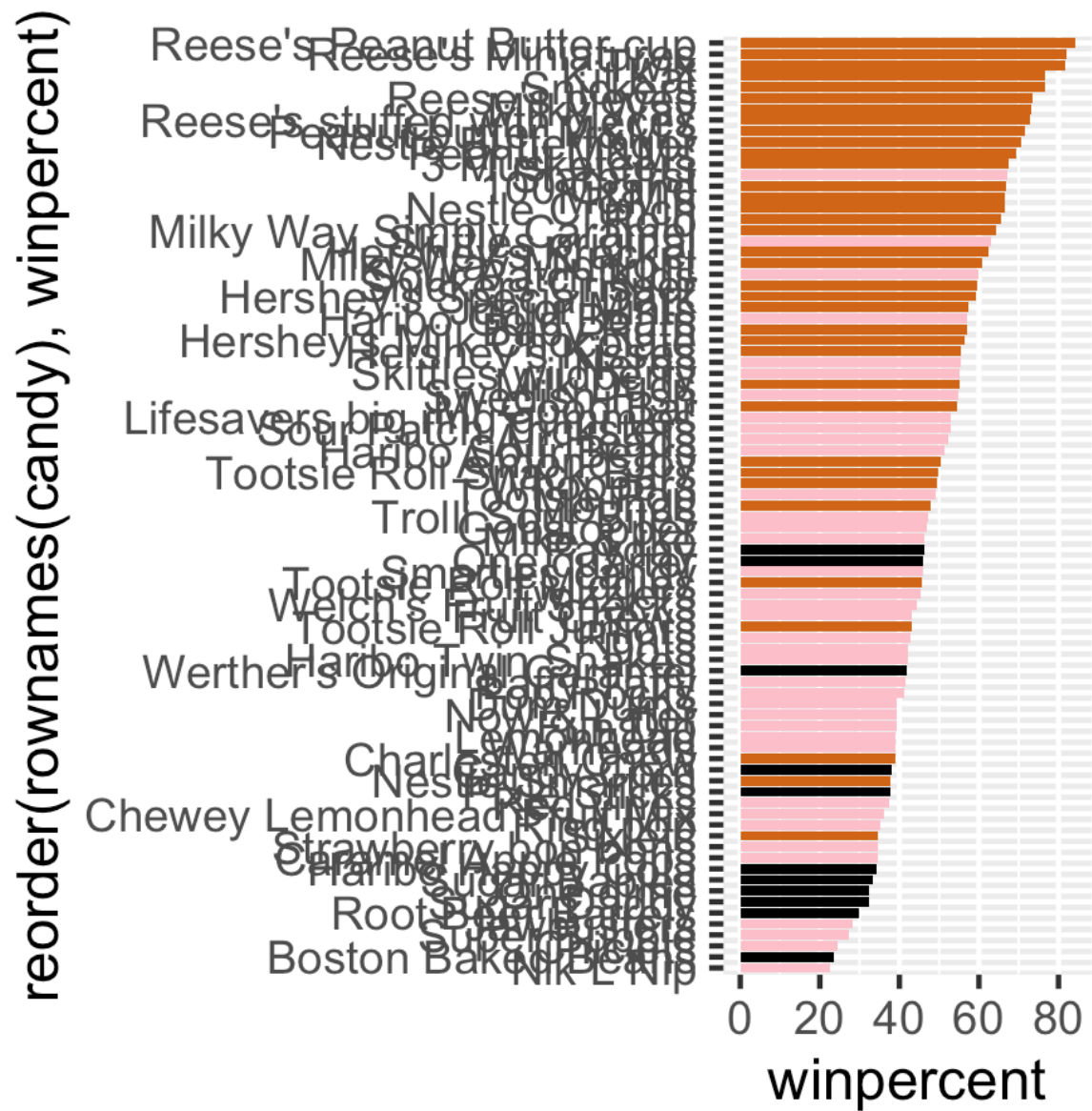
```
ggplot(candy) +
  aes(winpercent, reorder(rownames(candy), winpercent)) +
  geom_col()
```



Here, we want a custom color vector to color each bar the way we want - with `chocolate`, and `fruity` candy together, whether it is a bar or not.

```
my_cols=rep("black", nrow(candy))
my_cols[as.logical(candy$chocolate)] = "chocolate"
my_cols[as.logical(candy$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?

Starburst.

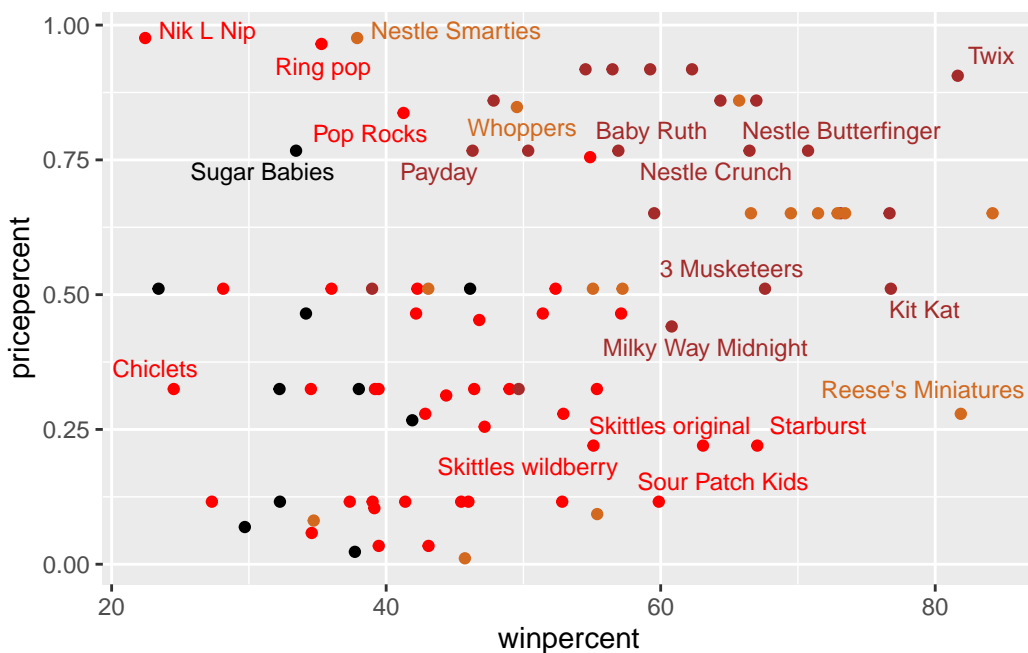
4. Taking a look at pricepercent

```
library(ggrepel)
```

```
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)] = "red"

ggplot(candy) +
  aes(winnerpercent, pricepercent, label=rownames(candy)) +
  geom_point(col=my_cols) +
  geom_text_repel(col=my_cols, size=3.3, max.overlaps = 5)
```

Warning: ggrepel: 65 unlabeled data points (too many overlaps). Consider increasing max.overlaps



Q19. Which candy type is the highest ranked in terms of winnerpercent 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?

```
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
Hershey's Krackel	0.918	62.28448
Hershey's Milk Chocolate	0.918	56.49050

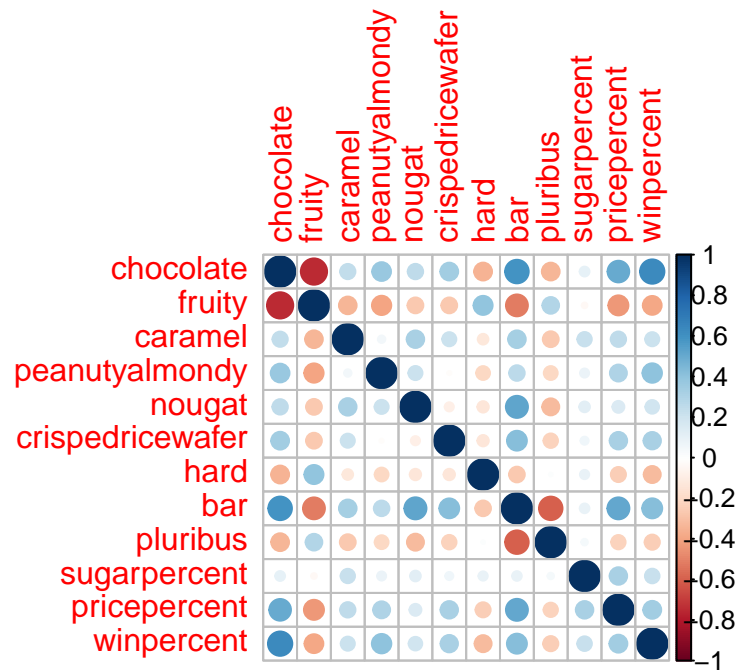
The least popular of the ones below happens to be the most expensive, which is the Nik L Nip.

##5. Exploring the correlation structure

```
library(corrplot)
```

corrplot 0.95 loaded

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



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

The ones that appear to be the most anti-correlated are chocolate-fruity or fruity-chocolate.

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

```
cij["chocolate","winpercent"]
```

```
[1] 0.6365167
```

```
cij["chocolate","bar"]
```

```
[1] 0.5974211
```

The most positively correlated variables would be chocolate and winpercent, as when compared to the only other circle that is of the same color or darker blue, it is higher.

##6. Principal Component Analysis

We need to be sure to scale our input `candy` data before PCA as we have the `winpercent` column on a different scale to all others in the dataset.

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

Importance of components:

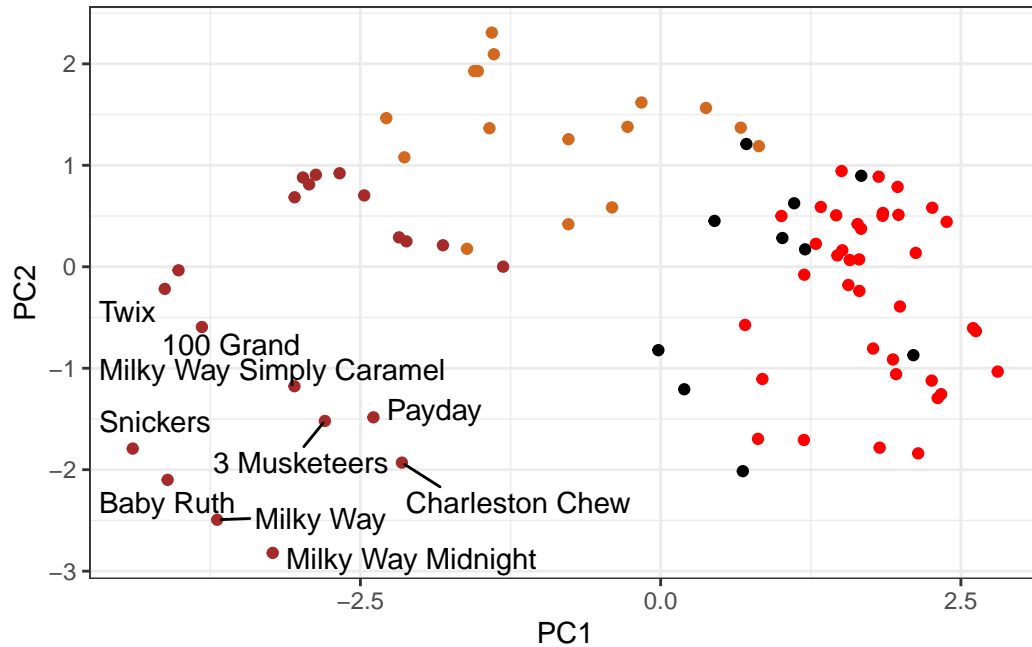
	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	2.0788	1.1378	1.1092	1.07533	0.9518	0.81923	0.81530
Proportion of Variance	0.3601	0.1079	0.1025	0.09636	0.0755	0.05593	0.05539
Cumulative Proportion	0.3601	0.4680	0.5705	0.66688	0.7424	0.79830	0.85369

	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

First main result figure is my “PCA plot”.

```
ggplot(pca$x) +
  aes(PC1, PC2, label=rownames(pca$x)) +
  geom_point(col=my_cols) +
  geom_text_repel(max.overlaps = 4) +
  theme_bw()
```

Warning: ggrepel: 75 unlabeled data points (too many overlaps). Consider increasing max.overlaps



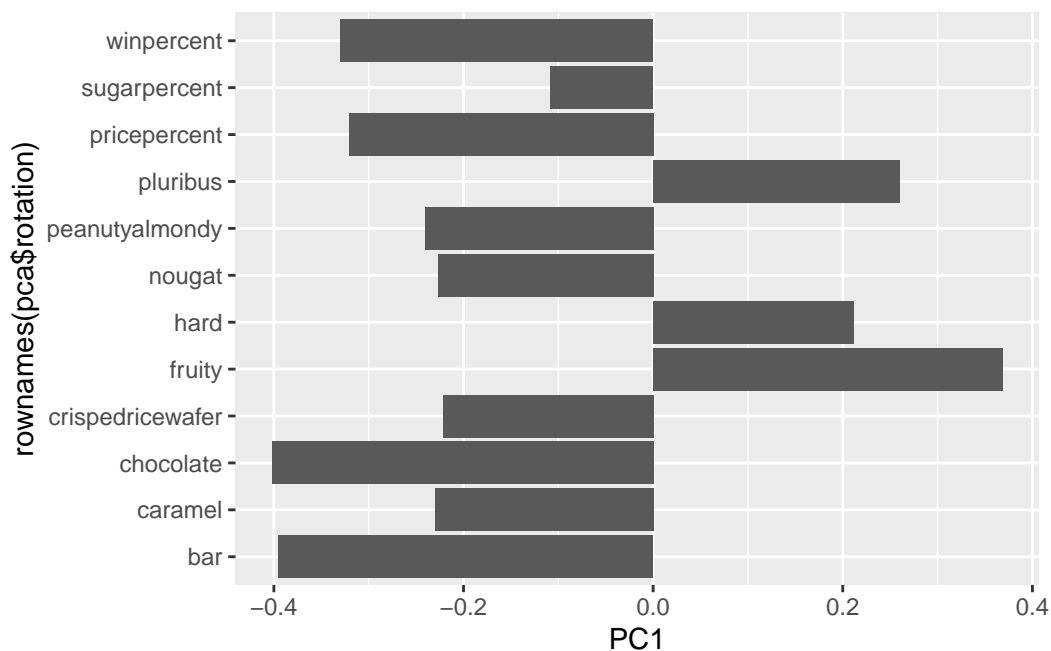
The second main PCA result is in the `pca$rotation`. We can plot this to generate a so-called “loadings” plot.

```
pca$rotation
```

	PC1	PC2	PC3	PC4	PC5
chocolate	-0.4019466	0.21404160	0.01601358	-0.016673032	0.066035846
fruity	0.3683883	-0.18304666	-0.13765612	-0.004479829	0.143535325
caramel	-0.2299709	-0.40349894	-0.13294166	-0.024889542	-0.507301501
peanutyalmondy	-0.2407155	0.22446919	0.18272802	0.466784287	0.399930245
nougat	-0.2268102	-0.47016599	0.33970244	0.299581403	-0.188852418
crispedricewafer	-0.2215182	0.09719527	-0.36485542	-0.605594730	0.034652316
hard	0.2111587	-0.43262603	-0.20295368	-0.032249660	0.574557816
bar	-0.3947433	-0.22255618	0.10696092	-0.186914549	0.077794806
pluribus	0.2600041	0.36920922	-0.26813772	0.287246604	-0.392796479
sugarpercent	-0.1083088	-0.23647379	-0.65509692	0.433896248	0.007469103
pricepercent	-0.3207361	0.05883628	-0.33048843	0.063557149	0.043358887
winpercent	-0.3298035	0.21115347	-0.13531766	0.117930997	0.168755073
	PC6	PC7	PC8	PC9	PC10
chocolate	-0.09018950	-0.08360642	-0.49084856	-0.151651568	0.107661356
fruity	-0.04266105	0.46147889	0.39805802	-0.001248306	0.362062502
caramel	-0.40346502	-0.44274741	0.26963447	0.019186442	0.229799010
peanutyalmondy	-0.09416259	-0.25710489	0.45771445	0.381068550	-0.145912362

nougat	0.09012643	0.36663902	-0.18793955	0.385278987	0.011323453
crispedricewafer	-0.09007640	0.13077042	0.13567736	0.511634999	-0.264810144
hard	-0.12767365	-0.31933477	-0.38881683	0.258154433	0.220779142
bar	0.25307332	0.24192992	-0.02982691	0.091872886	-0.003232321
pluribus	0.03184932	0.04066352	-0.28652547	0.529954405	0.199303452
sugarpercent	0.02737834	0.14721840	-0.04114076	-0.217685759	-0.488103337
pricepercent	0.62908570	-0.14308215	0.16722078	-0.048991557	0.507716043
winpercent	-0.56947283	0.40260385	-0.02936405	-0.124440117	0.358431235
	PC11	PC12			
chocolate	0.10045278	0.69784924			
fruity	0.17494902	0.50624242			
caramel	0.13515820	0.07548984			
peanutyalmondy	0.11244275	0.12972756			
nougat	-0.38954473	0.09223698			
crispedricewafer	-0.22615618	0.11727369			
hard	0.01342330	-0.10430092			
bar	0.74956878	-0.22010569			
pluribus	0.27971527	-0.06169246			
sugarpercent	0.05373286	0.04733985			
pricepercent	-0.26396582	-0.06698291			
winpercent	-0.11251626	-0.37693153			

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

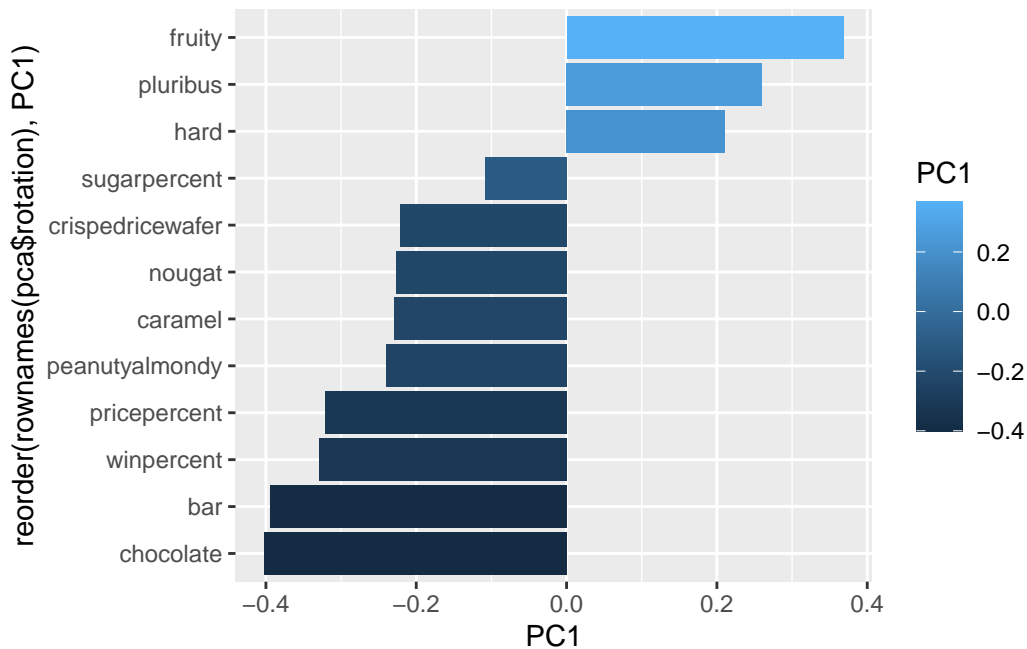


pca\$rotation

	PC1	PC2	PC3	PC4	PC5
chocolate	-0.4019466	0.21404160	0.01601358	-0.016673032	0.066035846
fruity	0.3683883	-0.18304666	-0.13765612	-0.004479829	0.143535325
caramel	-0.2299709	-0.40349894	-0.13294166	-0.024889542	-0.507301501
peanutyalmondy	-0.2407155	0.22446919	0.18272802	0.466784287	0.399930245
nougat	-0.2268102	-0.47016599	0.33970244	0.299581403	-0.188852418
crispedricewafer	-0.2215182	0.09719527	-0.36485542	-0.605594730	0.034652316
hard	0.2111587	-0.43262603	-0.20295368	-0.032249660	0.574557816
bar	-0.3947433	-0.22255618	0.10696092	-0.186914549	0.077794806
pluribus	0.2600041	0.36920922	-0.26813772	0.287246604	-0.392796479
sugarpercent	-0.1083088	-0.23647379	-0.65509692	0.433896248	0.007469103
pricepercent	-0.3207361	0.05883628	-0.33048843	0.063557149	0.043358887
winpercent	-0.3298035	0.21115347	-0.13531766	0.117930997	0.168755073
	PC6	PC7	PC8	PC9	PC10
chocolate	-0.09018950	-0.08360642	-0.49084856	-0.151651568	0.107661356
fruity	-0.04266105	0.46147889	0.39805802	-0.001248306	0.362062502
caramel	-0.40346502	-0.44274741	0.26963447	0.019186442	0.229799010
peanutyalmondy	-0.09416259	-0.25710489	0.45771445	0.381068550	-0.145912362
nougat	0.09012643	0.36663902	-0.18793955	0.385278987	0.011323453
crispedricewafer	-0.09007640	0.13077042	0.13567736	0.511634999	-0.264810144

hard	-0.12767365	-0.31933477	-0.38881683	0.258154433	0.220779142
bar	0.25307332	0.24192992	-0.02982691	0.091872886	-0.003232321
pluribus	0.03184932	0.04066352	-0.28652547	0.529954405	0.199303452
sugarpercent	0.02737834	0.14721840	-0.04114076	-0.217685759	-0.488103337
pricepercent	0.62908570	-0.14308215	0.16722078	-0.048991557	0.507716043
winpercent	-0.56947283	0.40260385	-0.02936405	-0.124440117	0.358431235
	PC11	PC12			
chocolate	0.10045278	0.69784924			
fruity	0.17494902	0.50624242			
caramel	0.13515820	0.07548984			
peanutyalmondy	0.11244275	0.12972756			
nougat	-0.38954473	0.09223698			
crispedricewafer	-0.22615618	0.11727369			
hard	0.01342330	-0.10430092			
bar	0.74956878	-0.22010569			
pluribus	0.27971527	-0.06169246			
sugarpercent	0.05373286	0.04733985			
pricepercent	-0.26396582	-0.06698291			
winpercent	-0.11251626	-0.37693153			

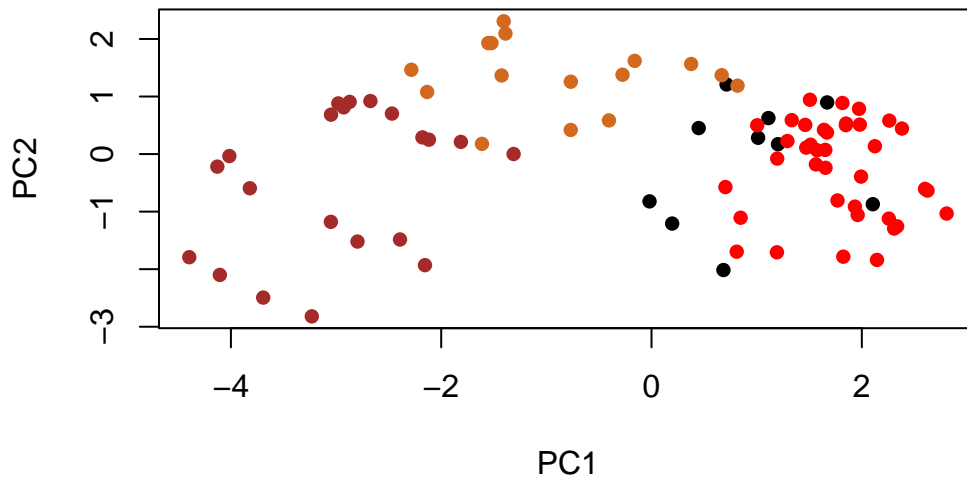
```
ggplot(pca$rotation) +
  aes(PC1, reorder(rownames(pca$rotation), PC1), fill=PC1) +
  geom_col()
```



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

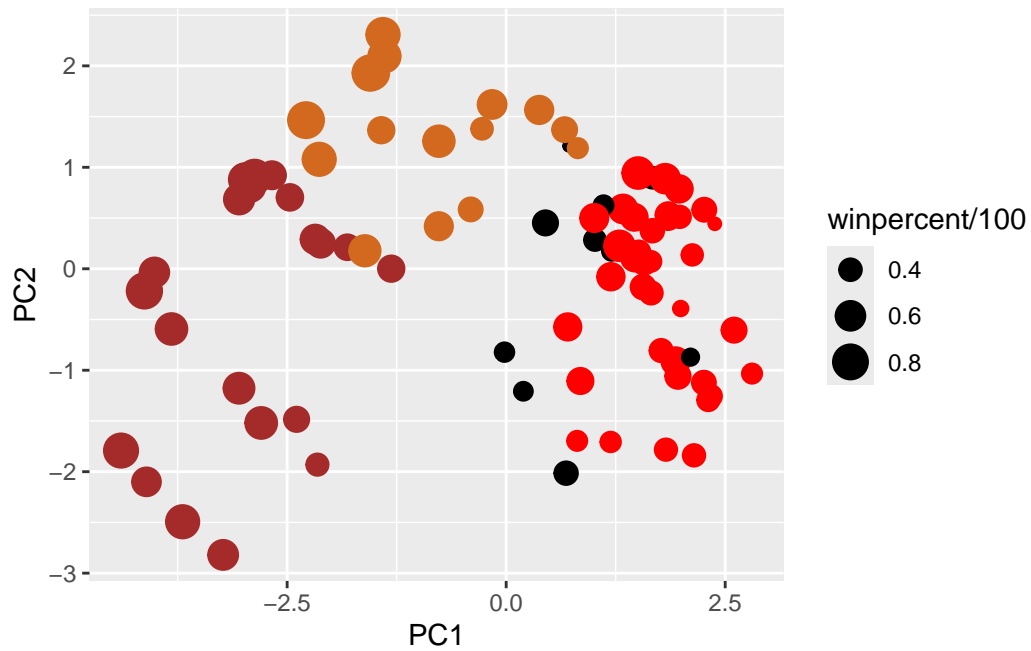
Fruity, hard, and pluribus.

```
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

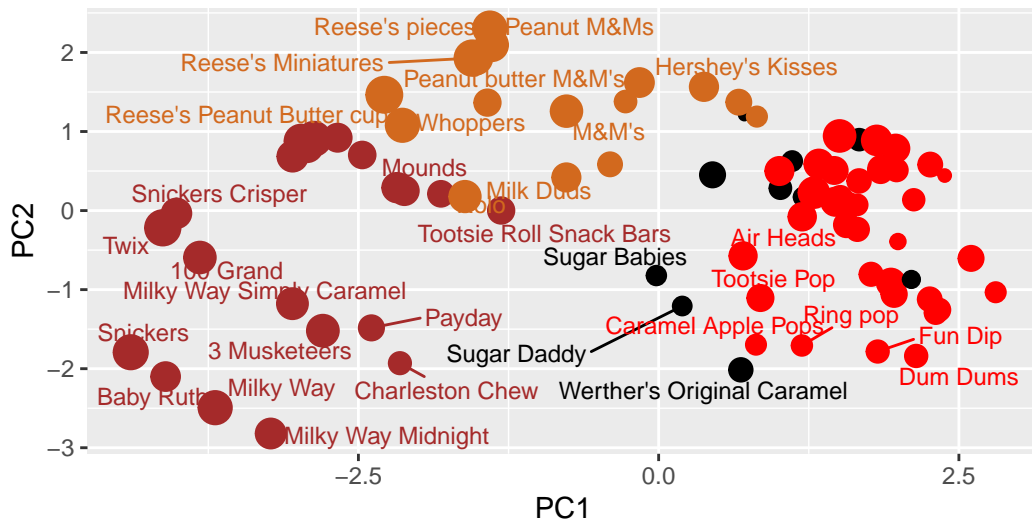


```
p + geom_text_repel(size=3.3, col=my_cols, max.overlaps = 10) +
  theme(legend.position = "none") +
  labs(title="Halloween Candy PCA Space",
        subtitle="Colored by type: chocolate bar (dark brown), chocolate other (light brown),",
        caption="Data from 538")
```

Warning: ggrepel: 53 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),



Data from 538