

Class08 - Halloween

Assael Madrigal (PID: A10179083)

load the file

```
candy_file <- "https://raw.githubusercontent.com/fivethirtyeight/data/master/candy-power-ratings.csv"

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

	chocolate	fruity	caramel	peanut	almond	nougat	crisped	rice	wafer
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
```

There are 85 different candies

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

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

```
[1] 38
```

```
#we can also look at which ones are fruity with logicals  
candy[ as.logical(candy$fruity),]
```

	chocolate	fruity	caramel	peanutyalmondy	nougat
Air Heads	0	1	0	0	0
Caramel Apple Pops	0	1	1	0	0
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 Sour Bears	0	1	0	0	0
Haribo Twin Snakes	0	1	0	0	0
Jawbusters	0	1	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
Mike & Ike	0	1	0	0	0
Nerds	0	1	0	0	0
Nik L Nip	0	1	0	0	0
Now & Later	0	1	0	0	0
Pop Rocks	0	1	0	0	0
Red vines	0	1	0	0	0
Ring pop	0	1	0	0	0
Runts	0	1	0	0	0
Skittles original	0	1	0	0	0
Skittles wildberry	0	1	0	0	0
Smarties candy	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
Strawberry bon bons	0	1	0	0	0
Super Bubble	0	1	0	0	0

Swedish Fish	0	1	0	0	0
Tootsie Pop	1	1	0	0	0
Trolli Sour Bites	0	1	0	0	0
Twizzlers	0	1	0	0	0
Warheads	0	1	0	0	0
Welch's Fruit Snacks	0	1	0	0	0
	crisped	rice	wafer	hard	bar pluribus sugarpercent
Air Heads		0	0	0	0 0.906
Caramel Apple Pops		0	0	0	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 Sour Bears		0	0	0	1 0.465
Haribo Twin Snakes		0	0	0	1 0.465
Jawbusters		0	1	0	1 0.093
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
Mike & Ike		0	0	0	1 0.872
Nerds		0	1	0	1 0.848
Nik L Nip		0	0	0	1 0.197
Now & Later		0	0	0	1 0.220
Pop Rocks		0	1	0	1 0.604
Red vines		0	0	0	1 0.581
Ring pop		0	1	0	0 0.732
Runts		0	1	0	1 0.872
Skittles original		0	0	0	1 0.941
Skittles wildberry		0	0	0	1 0.941
Smarties candy		0	1	0	1 0.267
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
Super Bubble		0	0	0	0 0.162
Swedish Fish		0	0	0	1 0.604
Tootsie Pop		0	1	0	0 0.604
Trolli Sour Bites		0	0	0	1 0.313
Twizzlers		0	0	0	0 0.220

Warheads	0	1	0	0	0.093
Welch's Fruit Snacks	0	0	0	1	0.313
	price	percent	win	percent	
Air Heads	0.511	52.34	146		
Caramel Apple Pops	0.325	34.51	768		
Chewey Lemonhead Fruit Mix	0.511	36.01	763		
Chiclets	0.325	24.52	499		
Dots	0.511	42.27	208		
Dum Dums	0.034	39.46	056		
Fruit Chews	0.034	43.08	892		
Fun Dip	0.325	39.18	550		
Gobstopper	0.453	46.78	335		
Haribo Gold Bears	0.465	57.11	974		
Haribo Sour Bears	0.465	51.41	243		
Haribo Twin Snakes	0.465	42.17	877		
Jawbusters	0.511	28.12	744		
Laffy Taffy	0.116	41.38	956		
Lemonhead	0.104	39.14	106		
Lifesavers big ring gummies	0.279	52.91	139		
Mike & Ike	0.325	46.41	172		
Nerds	0.325	55.35	405		
Nik L Nip	0.976	22.44	534		
Now & Later	0.325	39.44	680		
Pop Rocks	0.837	41.26	551		
Red vines	0.116	37.34	852		
Ring pop	0.965	35.29	076		
Runts	0.279	42.84	914		
Skittles original	0.220	63.08	514		
Skittles wildberry	0.220	55.10	370		
Smarties candy	0.116	45.99	583		
Sour Patch Kids	0.116	59.86	400		
Sour Patch Tricksters	0.116	52.82	595		
Starburst	0.220	67.03	763		
Strawberry bon bons	0.058	34.57	899		
Super Bubble	0.116	27.30	386		
Swedish Fish	0.755	54.86	111		
Tootsie Pop	0.325	48.98	265		
Trolli Sour Bites	0.255	47.17	323		
Twizzlers	0.116	45.46	628		
Warheads	0.116	39.01	190		
Welch's Fruit Snacks	0.313	44.37	552		

There are 38 types of fruity candy

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

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

```
[1] 76.67378
```

Snickers has a win percent of 76%

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

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

```
[1] 76.7686
```

Kit kat also has a winpercent of 76%

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

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

```
[1] 49.6535
```

Tootsie rolls has a 49% winpercent

```
#library("skimr")  
#this way i don't have to load the whole package, just the function  
#skimr::skim  
#skim_candy <- skim(candy)  
#skim_candy
```

Q6. Is there any variable/column that looks to be on a different scale to the majority of the other columns in the dataset?

All of the data has a value between 0 and 1 but the winrate is between 0 and 100, so it would dominate the PCA. So we need to **scale** it

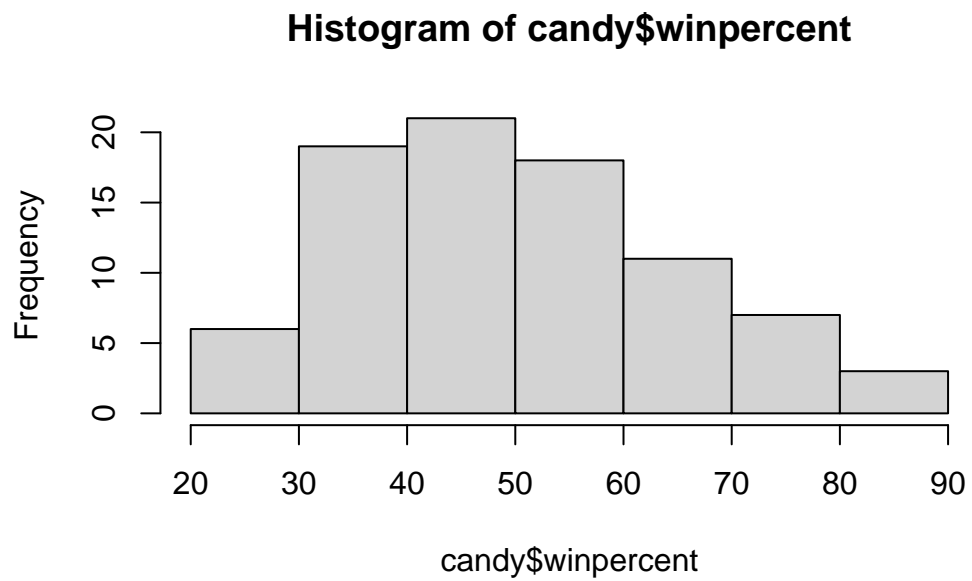
Q7. What do you think a zero and one represent for the candy\$chocolate column?

0 means that the candy has no chocolate, 1 it means it has it

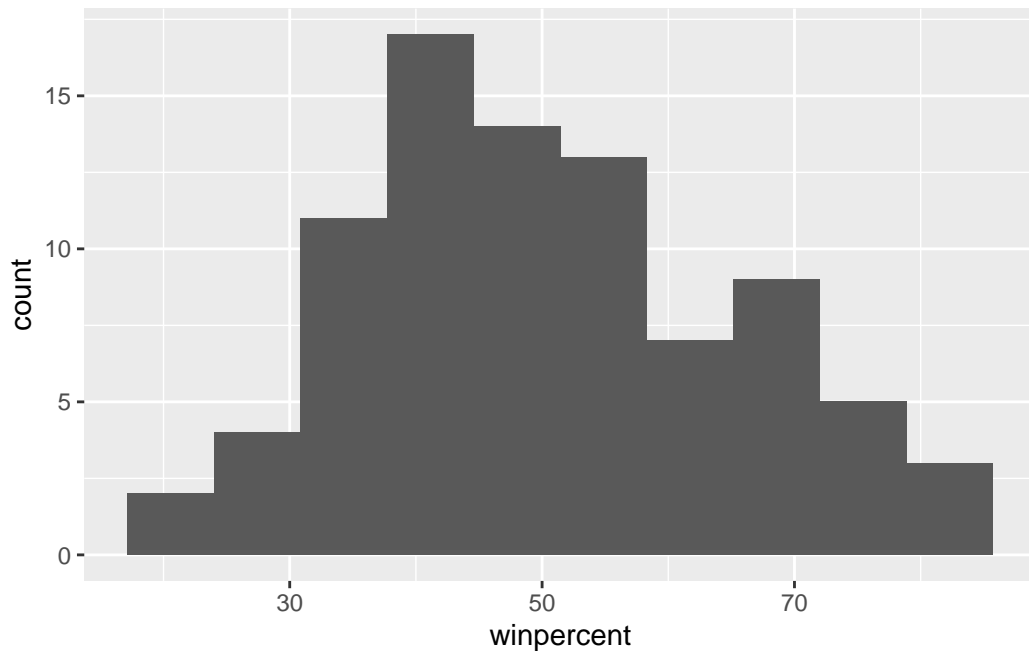
Q8. Plot a histogram of winpercent values

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

```
library(ggplot2)
```



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



Q9. Is the distribution of winpercent values symmetrical?

Not perfectly symmetrical, but very close to it so it does not deviate too much and the center looks very close to the median.

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

```
median(candy$winpercent)
```

```
[1] 47.82975
```

The center of distribution is slightly below 50%

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

```
mean(candy[as.logical(candy$chocolate),"winpercent"])
```

```
[1] 60.92153
```

```
mean(candy[as.logical(candy$fruity),"winpercent"])
```

```
[1] 44.11974
```

```
mean(candy[as.logical(candy$chocolate),"winpercent"]) > mean(candy[as.logical(candy$fruity),"winpercent"])
```

```
[1] TRUE
```

Yes it is, chocolate is 61% and fruity is 44%

Q12. Is this difference statistically significant?

```
t.test(candy[as.logical(candy$chocolate),"winpercent"], candy[as.logical(candy$fruity),"winpercent"], var.equal=FALSE)
```

Welch Two Sample t-test

```
data: candy[as.logical(candy$chocolate), "winpercent"] and candy[as.logical(candy$fruity), "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 is significant because the p-value is less than 0.05

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

```
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
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
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
Chewy 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
Chewy 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	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

The worst 5 are Nik, Boston, chiclets, super and jawbusters

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

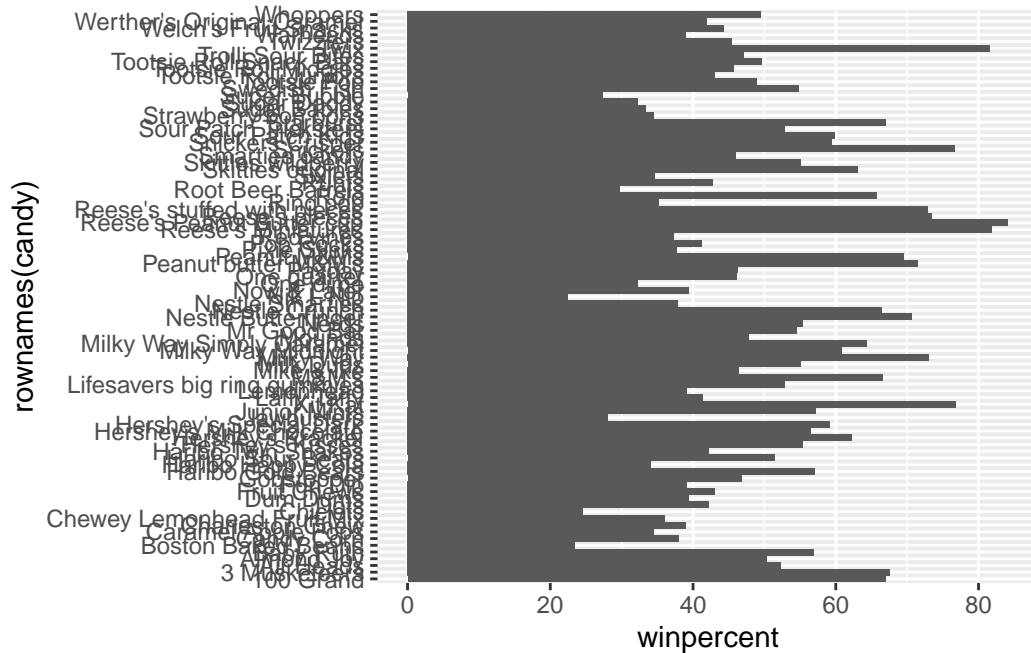
```
head(candy[order(candy$winpercent, decreasing=TRUE),], n=5)
```

	chocolate	fruity	caramel	peanut	almond	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
	crisped	rice	wafer	hard	bar	pluribus
Reese's Peanut Butter cup		0	0	0		0.720
Reese's Miniatures		0	0	0		0.034
Twix		1	0	1		0.546
Kit Kat		1	0	1		0.313
Snickers		0	0	1		0.546
	price	percent	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			

The top 5 favorites are reeses, reeses, twix, kit kat and snickers

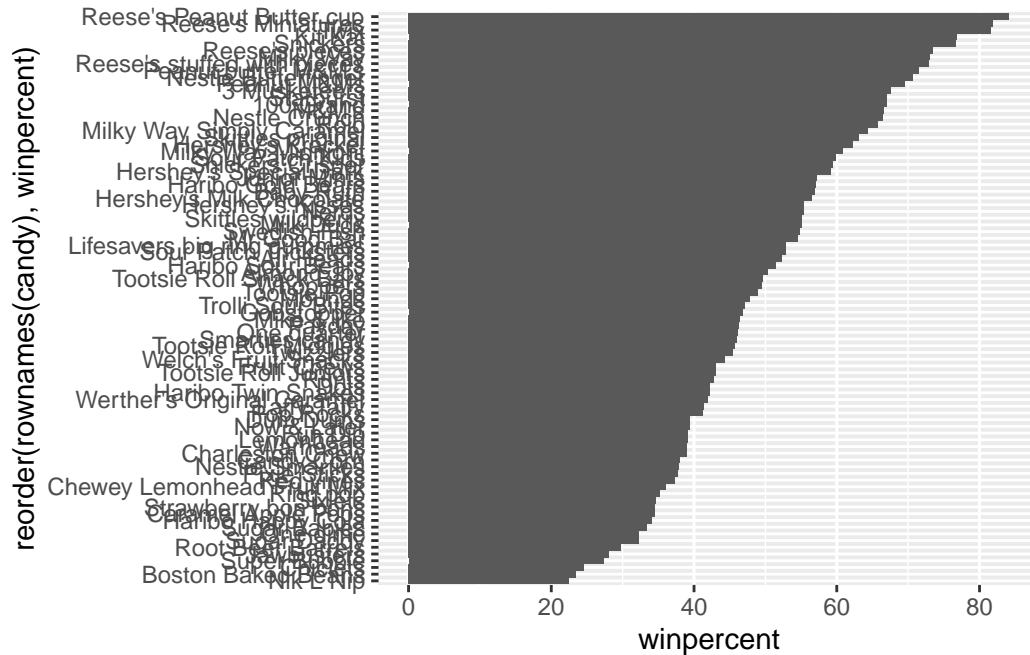
Q15. Make a first barplot of candy ranking based on winpercent values.

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



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

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

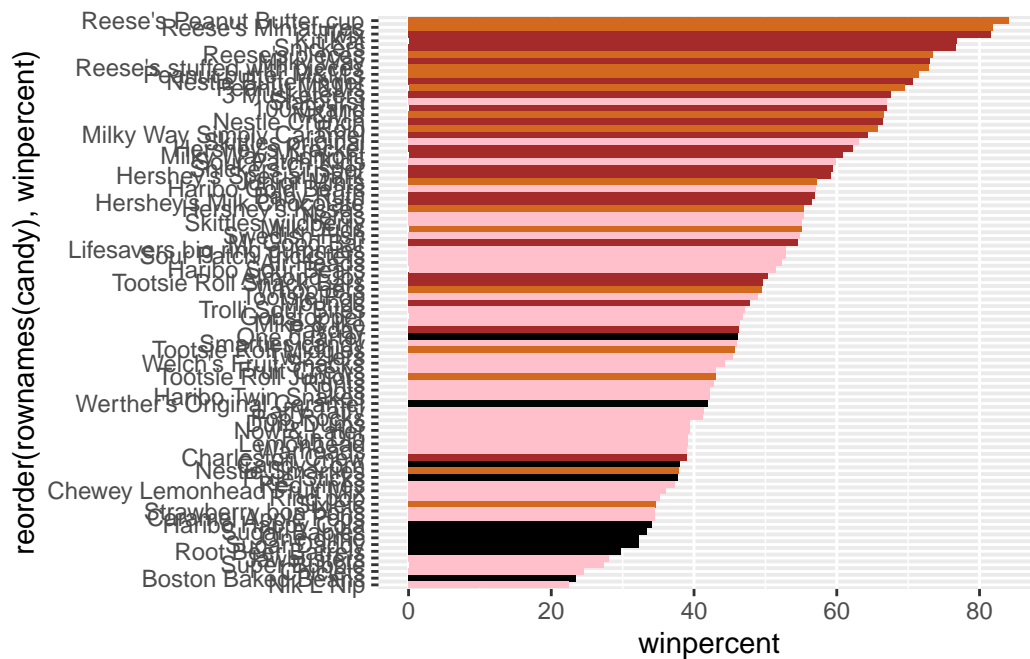
Color a vector by if it has chocolates

```
#rep() just makes a repeat of what we say plus the number of times
mycols <- rep("gray", nrow(candy))

mycols[ as.logical(candy$chocolate)] <- "blue"

#applying this to candy
my_cols=rep("black", nrow(candy))
my_cols[as.logical(candy$chocolate)] = "chocolate"
my_cols[as.logical(candy$bar)] = "brown"
my_cols[as.logical(candy$fruity)] = "pink"

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



Q17. What is the worst ranked chocolate candy?

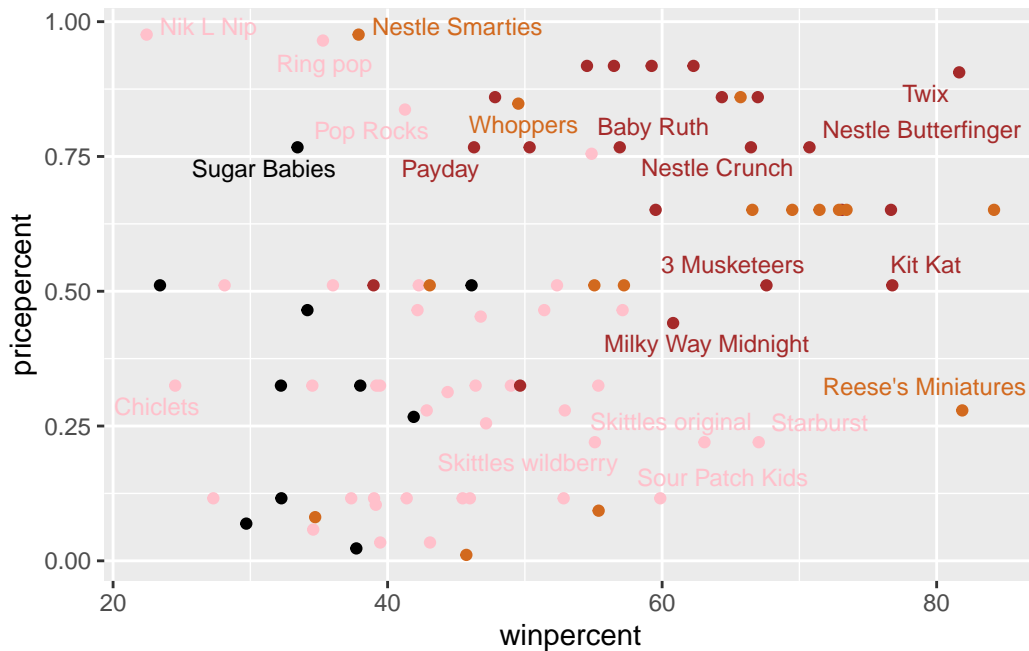
Sixlets are ranked the worst chocolate > Q18. What is the best ranked fruity candy?

Starburst is ranked the best 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: 65 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?

```
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_mutated <- candy %>%
  mutate(bang_per_buck = pricepercent/winpercent)%>%
  arrange(bang_per_buck)
```

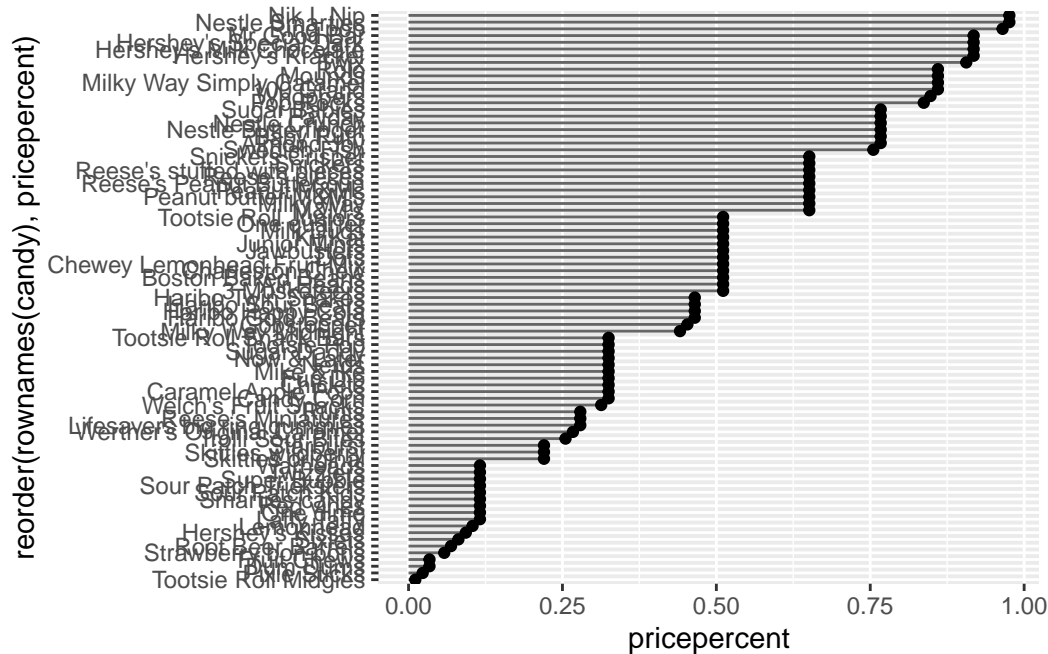
Tootsie Roll is the best candy for bang per buck

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

Top most expensive are NikLNip, Ring pop, Nestle Smarties, Mily way and Simply caramel of these the least popular is NikLNip

Optional 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()`.

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

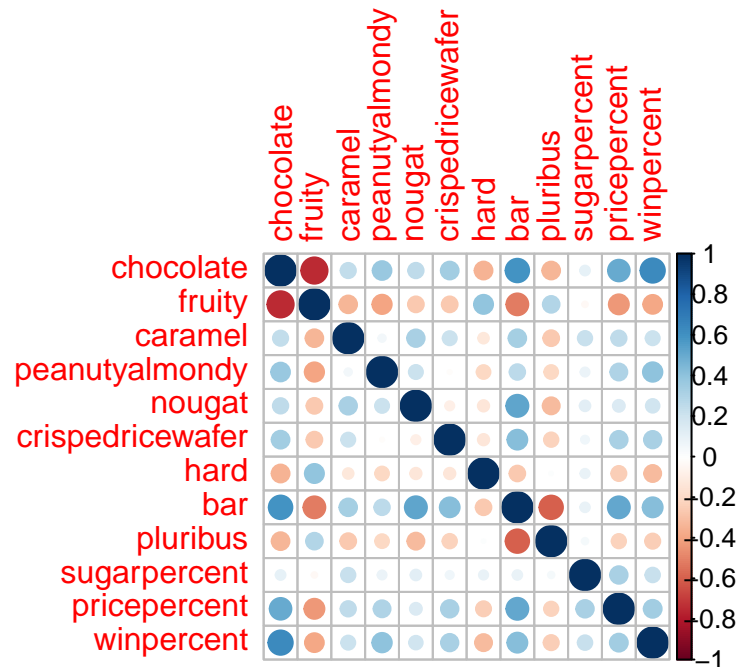


```
library(corrplot)
```

corrplot 0.92 loaded

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

```
corrplot(cij)
```



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

Fruit and chocolate are the most anti-correlated > Q23. Similarly, what two variables are most positively correlated?

The most correlated is winpercent and chocolate with a correlation of 0.63

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

```
summary(pca)
```

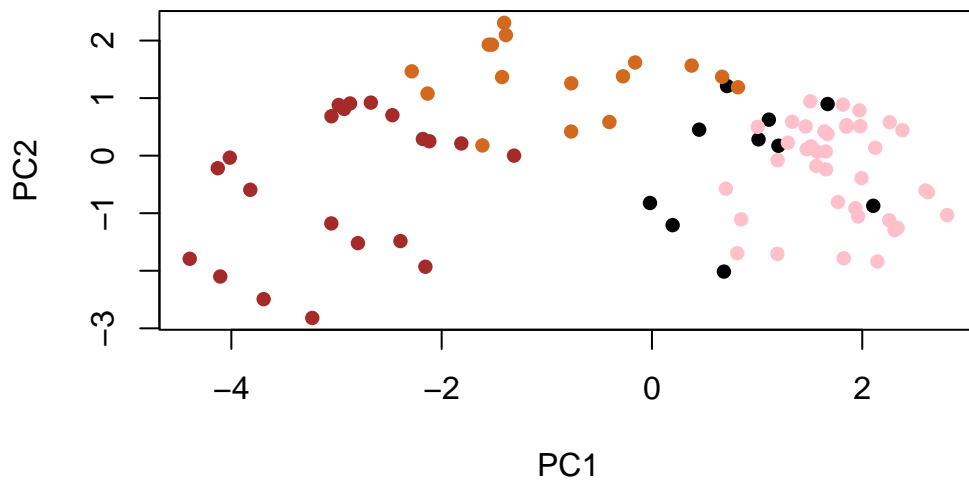
Importance of components:

	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

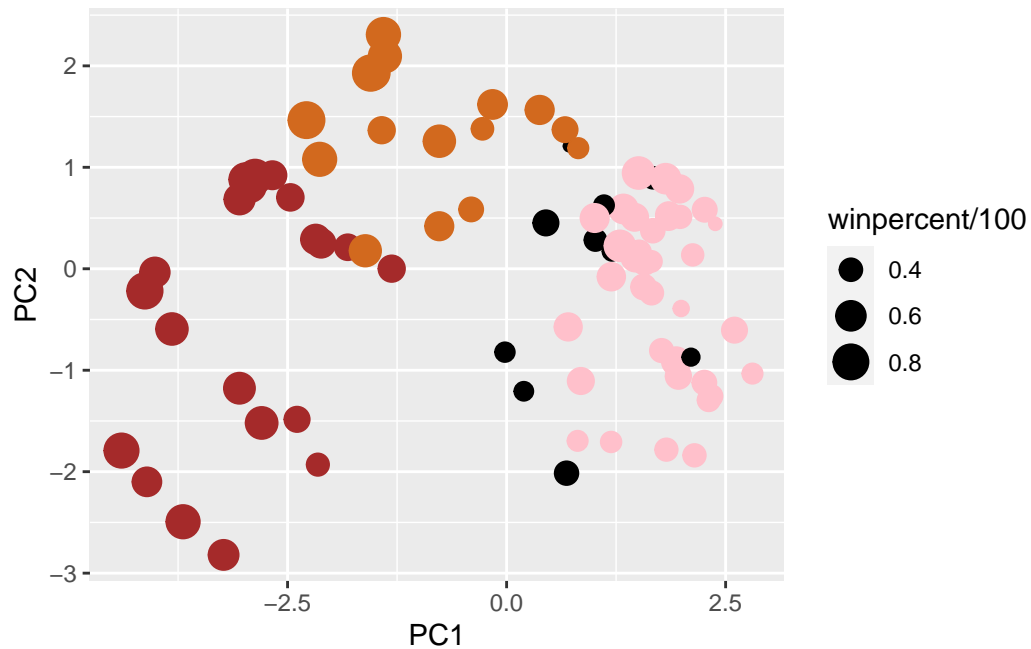
Proportion of Variance	0.04629	0.03833	0.03239	0.01611	0.01317
Cumulative Proportion	0.89998	0.93832	0.97071	0.98683	1.00000

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



```
# 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)
```

p



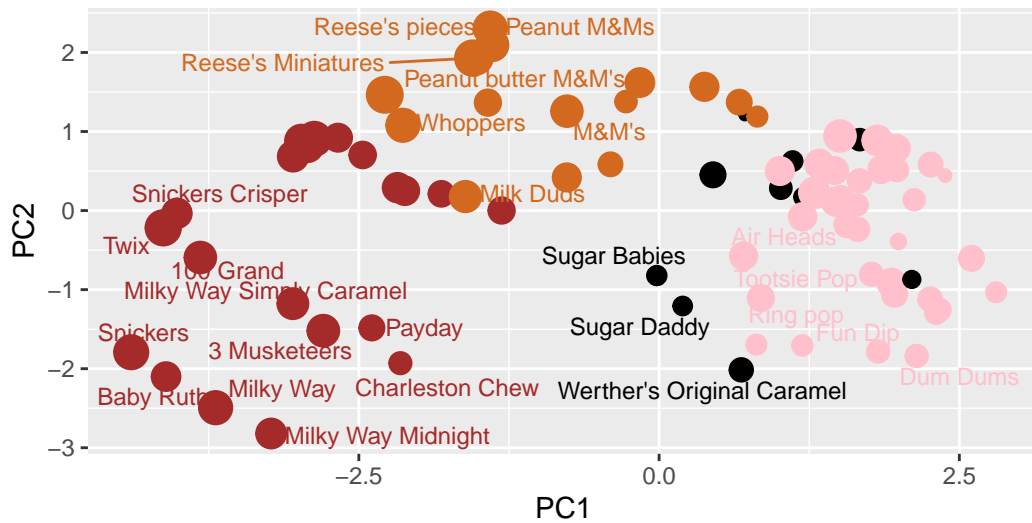
```
library(ggrepel)

p + geom_text_repel(size=3.3, col=my_cols, max.overlaps = 7) +
  theme(legend.position = "none") +
  labs(title="Halloween Candy PCA Space",
        subtitle="Colored by type: chocolate bar (dark brown), chocolate other (light brown)",
        caption="Data from 538")
```

Warning: ggrepel: 59 unlabeled data points (too many overlaps). Consider increasing max.overlaps

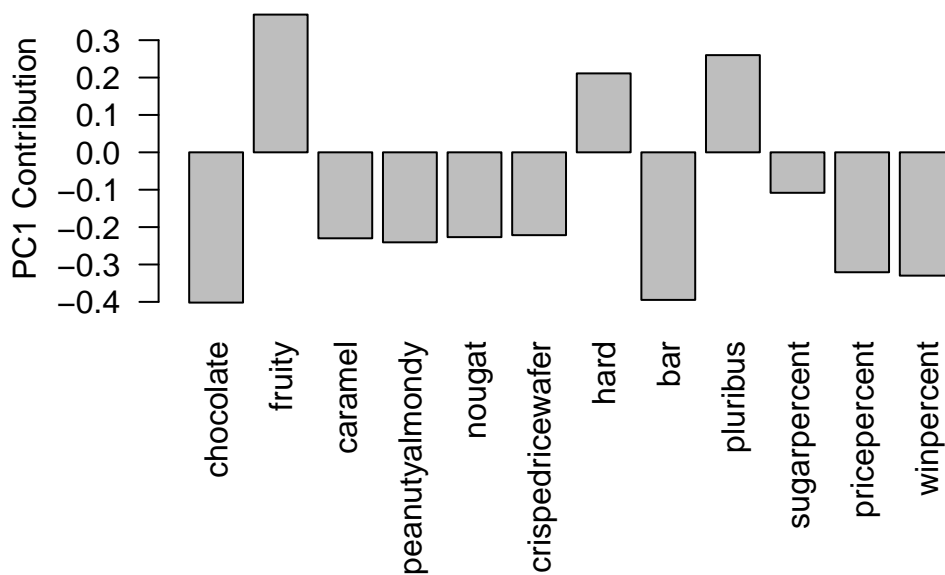
Halloween Candy PCA Space

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



Data from 538

```
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 most contribution for PC1 is fruity, then pluribus then hard