

# class10 halloween mini project

Ryan Bench (PID:A69038034)

```
candy_file <- "https://raw.githubusercontent.com/fivethirtyeight/data/master/candy-power-rankings.csv"
candy = read.csv(candy_file, row.names=1)
head(candy)
```

	chocolate	fruity	caramel	peanut	yalmond	nougat	crispedrice	wafers
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		

```
dim(candy)
```

[1] 85 12

```
nrow(candy)
```

[1] 85

```
sum(candy)
```

```
[1] 4554.465
```

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

There are 12 candy types in this dataset, with 85 different individual candies.

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

```
[1] 38
```

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

There are 38 fruity candy types in the dataset.

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

```
[1] 81.64291
```

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

```
[1] 73.09956
```

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

My favorite candy in the dataset is a Milky Way bar, and the winpercent value is 73.09956.

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

```
[1] 76.7686
```

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

The winpercent value is 76.7686.

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

```
[1] 49.6535
```

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

The winpercent value for Tootsie Roll Snack Bars is 49.6535.

```
#skimr would not render
```

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 scale is on 0-100 scale, the rest are on 0-1.

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

I believe 0 means the candy is not chocolate, but a 1 means it is chocolate.

```
library(dplyr)
```

```
candy |> nrow()
```

```
[1] 85
```

```
round( mean(candy$winpercent), 4)
```

```
[1] 50.3168
```

```
candy |> select(winpercent)
```

	winpercent
100 Grand	66.97173
3 Musketeers	67.60294
One dime	32.26109
One quarter	46.11650
Air Heads	52.34146
Almond Joy	50.34755
Baby Ruth	56.91455
Boston Baked Beans	23.41782
Candy Corn	38.01096
Caramel Apple Pops	34.51768
Charleston Chew	38.97504
Chewey Lemonhead Fruit Mix	36.01763
Chiclets	24.52499
Dots	42.27208
Dum Dums	39.46056
Fruit Chews	43.08892
Fun Dip	39.18550

Gobstopper	46.78335
Haribo Gold Bears	57.11974
Haribo Happy Cola	34.15896
Haribo Sour Bears	51.41243
Haribo Twin Snakes	42.17877
Hershey's Kisses	55.37545
Hershey's Krackel	62.28448
Hershey's Milk Chocolate	56.49050
Hershey's Special Dark	59.23612
Jawbusters	28.12744
Junior Mints	57.21925
Kit Kat	76.76860
Laffy Taffy	41.38956
Lemonhead	39.14106
Lifesavers big ring gummies	52.91139
Peanut butter M&M's	71.46505
M&M's	66.57458
Mike & Ike	46.41172
Milk Duds	55.06407
Milky Way	73.09956
Milky Way Midnight	60.80070
Milky Way Simply Caramel	64.35334
Mounds	47.82975
Mr Good Bar	54.52645
Nerds	55.35405
Nestle Butterfinger	70.73564
Nestle Crunch	66.47068
Nik L Nip	22.44534
Now & Later	39.44680
Payday	46.29660
Peanut M&Ms	69.48379
Pixie Sticks	37.72234
Pop Rocks	41.26551
Red vines	37.34852
Reese's Miniatures	81.86626
Reese's Peanut Butter cup	84.18029
Reese's pieces	73.43499
Reese's stuffed with pieces	72.88790
Ring pop	35.29076
Rolo	65.71629
Root Beer Barrels	29.70369
Runts	42.84914
Sixlets	34.72200

Skittles original	63.08514
Skittles wildberry	55.10370
Nestle Smarties	37.88719
Smarties candy	45.99583
Snickers	76.67378
Snickers Crisper	59.52925
Sour Patch Kids	59.86400
Sour Patch Tricksters	52.82595
Starburst	67.03763
Strawberry bon bons	34.57899
Sugar Babies	33.43755
Sugar Daddy	32.23100
Super Bubble	27.30386
Swedish Fish	54.86111
Tootsie Pop	48.98265
Tootsie Roll Juniors	43.06890
Tootsie Roll Midgies	45.73675
Tootsie Roll Snack Bars	49.65350
Trolli Sour Bites	47.17323
Twix	81.64291
Twizzlers	45.46628
Warheads	39.01190
Welch's Fruit Snacks	44.37552
Werther's Original Caramel	41.90431
Whoppers	49.52411

```
win <- candy$winpercent
win.mean <- mean(win)
round(win.mean)
```

[1] 50

```
candy %>% select(winpercent)
```

	winpercent
100 Grand	66.97173
3 Musketeers	67.60294
One dime	32.26109
One quarter	46.11650
Air Heads	52.34146
Almond Joy	50.34755

Baby Ruth	56.91455
Boston Baked Beans	23.41782
Candy Corn	38.01096
Caramel Apple Pops	34.51768
Charleston Chew	38.97504
Chewey Lemonhead Fruit Mix	36.01763
Chiclets	24.52499
Dots	42.27208
Dum Dums	39.46056
Fruit Chews	43.08892
Fun Dip	39.18550
Gobstopper	46.78335
Haribo Gold Bears	57.11974
Haribo Happy Cola	34.15896
Haribo Sour Bears	51.41243
Haribo Twin Snakes	42.17877
Hershey's Kisses	55.37545
Hershey's Krackel	62.28448
Hershey's Milk Chocolate	56.49050
Hershey's Special Dark	59.23612
Jawbusters	28.12744
Junior Mints	57.21925
Kit Kat	76.76860
Laffy Taffy	41.38956
Lemonhead	39.14106
Lifesavers big ring gummies	52.91139
Peanut butter M&M's	71.46505
M&M's	66.57458
Mike & Ike	46.41172
Milk Duds	55.06407
Milky Way	73.09956
Milky Way Midnight	60.80070
Milky Way Simply Caramel	64.35334
Mounds	47.82975
Mr Good Bar	54.52645
Nerds	55.35405
Nestle Butterfinger	70.73564
Nestle Crunch	66.47068
Nik L Nip	22.44534
Now & Later	39.44680
Payday	46.29660
Peanut M&Ms	69.48379
Pixie Sticks	37.72234

Pop Rocks	41.26551
Red vines	37.34852
Reese's Miniatures	81.86626
Reese's Peanut Butter cup	84.18029
Reese's pieces	73.43499
Reese's stuffed with pieces	72.88790
Ring pop	35.29076
Rolo	65.71629
Root Beer Barrels	29.70369
Runts	42.84914
Sixlets	34.72200
Skittles original	63.08514
Skittles wildberry	55.10370
Nestle Smarties	37.88719
Smarties candy	45.99583
Snickers	76.67378
Snickers Crisper	59.52925
Sour Patch Kids	59.86400
Sour Patch Tricksters	52.82595
Starburst	67.03763
Strawberry bon bons	34.57899
Sugar Babies	33.43755
Sugar Daddy	32.23100
Super Bubble	27.30386
Swedish Fish	54.86111
Tootsie Pop	48.98265
Tootsie Roll Juniors	43.06890
Tootsie Roll Midgies	45.73675
Tootsie Roll Snack Bars	49.65350
Trolli Sour Bites	47.17323
Twix	81.64291
Twizzlers	45.46628
Warheads	39.01190
Welch's Fruit Snacks	44.37552
Werther's Original Caramel	41.90431
Whoppers	49.52411

```
candy |> select(winpercent)
```

	winpercent
100 Grand	66.97173
3 Musketeers	67.60294

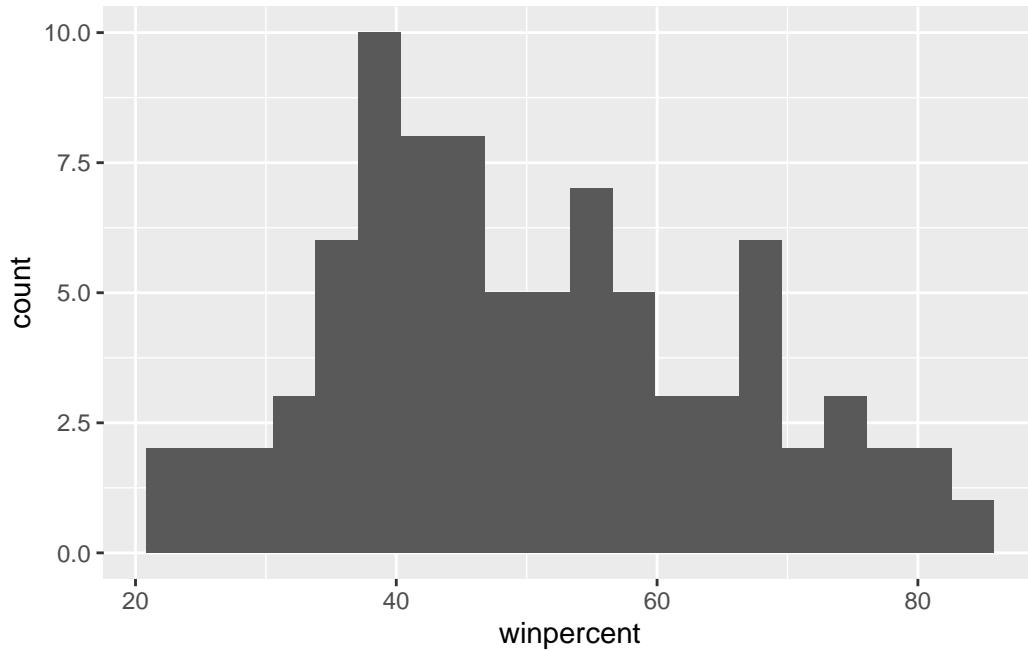
One dime	32.26109
One quarter	46.11650
Air Heads	52.34146
Almond Joy	50.34755
Baby Ruth	56.91455
Boston Baked Beans	23.41782
Candy Corn	38.01096
Caramel Apple Pops	34.51768
Charleston Chew	38.97504
Chewey Lemonhead Fruit Mix	36.01763
Chiclets	24.52499
Dots	42.27208
Dum Dums	39.46056
Fruit Chews	43.08892
Fun Dip	39.18550
Gobstopper	46.78335
Haribo Gold Bears	57.11974
Haribo Happy Cola	34.15896
Haribo Sour Bears	51.41243
Haribo Twin Snakes	42.17877
Hershey's Kisses	55.37545
Hershey's Krackel	62.28448
Hershey's Milk Chocolate	56.49050
Hershey's Special Dark	59.23612
Jawbusters	28.12744
Junior Mints	57.21925
Kit Kat	76.76860
Laffy Taffy	41.38956
Lemonhead	39.14106
Lifesavers big ring gummies	52.91139
Peanut butter M&M's	71.46505
M&M's	66.57458
Mike & Ike	46.41172
Milk Duds	55.06407
Milky Way	73.09956
Milky Way Midnight	60.80070
Milky Way Simply Caramel	64.35334
Mounds	47.82975
Mr Good Bar	54.52645
Nerds	55.35405
Nestle Butterfinger	70.73564
Nestle Crunch	66.47068
Nik L Nip	22.44534

Now & Later	39.44680
Payday	46.29660
Peanut M&Ms	69.48379
Pixie Sticks	37.72234
Pop Rocks	41.26551
Red vines	37.34852
Reese's Miniatures	81.86626
Reese's Peanut Butter cup	84.18029
Reese's pieces	73.43499
Reese's stuffed with pieces	72.88790
Ring pop	35.29076
Rolo	65.71629
Root Beer Barrels	29.70369
Runts	42.84914
Sixlets	34.72200
Skittles original	63.08514
Skittles wildberry	55.10370
Nestle Smarties	37.88719
Smarties candy	45.99583
Snickers	76.67378
Snickers Crisper	59.52925
Sour Patch Kids	59.86400
Sour Patch Tricksters	52.82595
Starburst	67.03763
Strawberry bon bons	34.57899
Sugar Babies	33.43755
Sugar Daddy	32.23100
Super Bubble	27.30386
Swedish Fish	54.86111
Tootsie Pop	48.98265
Tootsie Roll Juniors	43.06890
Tootsie Roll Midgies	45.73675
Tootsie Roll Snack Bars	49.65350
Trolli Sour Bites	47.17323
Twix	81.64291
Twizzlers	45.46628
Warheads	39.01190
Welch's Fruit Snacks	44.37552
Werther's Original Caramel	41.90431
Whoppers	49.52411

```
#These are the same
```

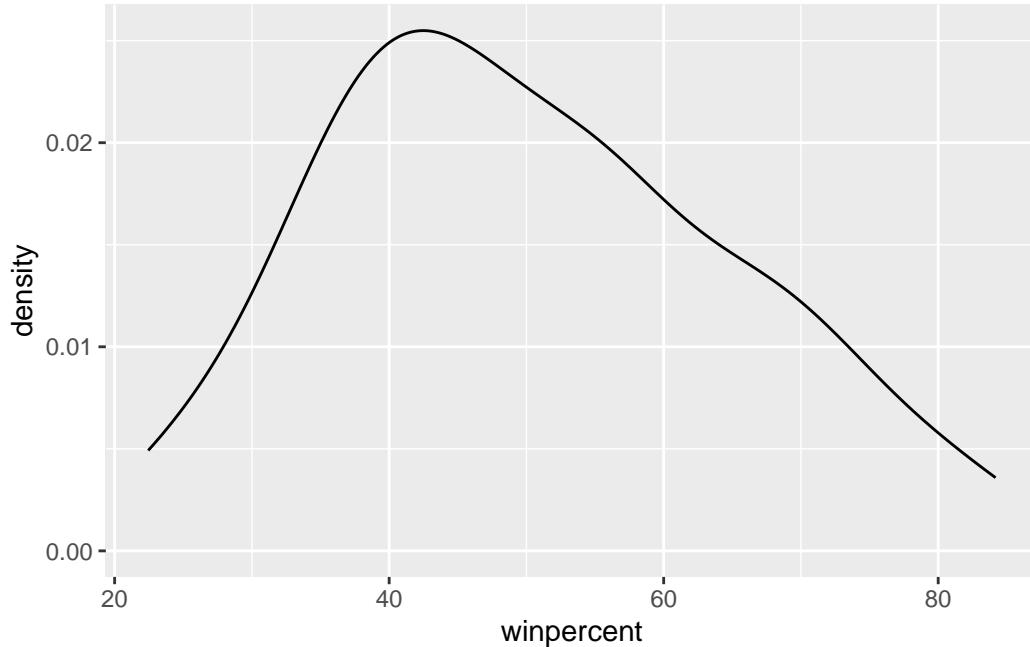
Q8. Plot a histogram of winpercent values

```
library(ggplot2)
ggplot(candy) + aes(winpercent) + geom_histogram(bins = 20)
```



Q9. Is the distribution of winpercent values symmetrical?

```
ggplot(candy) + aes(winpercent) + geom_density()
```



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

The center is below 50% according to the plots. According to the mean it is 50%.

```
mean(candy$winpercent)
```

```
[1] 50.31676
```

```
summary(candy$winpercent)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
22.45	39.14	47.83	50.32	59.86	84.18

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

```
#1. Find all chocolate candy in the dataset
choc inds <- as.logical(candy$chocolate)
choc candy <- candy[choc inds,]
choc candy
```

	chocolate	fruity	caramel	peanuty	almondy	nougat
100 Grand	1	0	1	0	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	
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	
Junior Mints	1	0	0	0	0	
Kit Kat	1	0	0	0	0	
Peanut butter M&M's	1	0	0	1	0	
M&M's	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	
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	
Rolo	1	0	1	0	0	
Sixlets	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	rice	wafer	hard	bar	pluribus sugarpercent
100 Grand	1	0	1	0	0	0.732
3 Musketeers	0	0	1	0	0	0.604
Almond Joy	0	0	1	0	0	0.465
Baby Ruth	0	0	1	0	0	0.604

Charleston Chew	0	0	1	0	0.604
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
Junior Mints	0	0	0	1	0.197
Kit Kat	1	0	1	0	0.313
Peanut butter M&M's	0	0	0	1	0.825
M&M's	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
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
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
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

Junior Mints	0.511	57.21925
Kit Kat	0.511	76.76860
Peanut butter M&M's	0.651	71.46505
M&M's	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
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
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

```
#2. Extract the `winpercent` values
choc.win <- choc.candy$winpercent
#3. Find the mean of these values
choc.mean <- mean(choc.win)
choc.mean
```

[1] 60.92153

```
#4-6. Do the same for fruity candy
fruity inds <- as.logical(candy$fruity)
fruity.candy <- candy[fruity.inds, ]
fruity.candy
```

	chocolate	fruity	caramel	peanuty	almondy	nougat
Air Heads	0	1	0	0	0	0
Caramel Apple Pops	0	1	1	0	0	0
Chewey Lemonhead Fruit Mix	0	1	0	0	0	0
Chiclets	0	1	0	0	0	0
Dots	0	1	0	0	0	0
Dum Dums	0	1	0	0	0	0
Fruit Chews	0	1	0	0	0	0
Fun Dip	0	1	0	0	0	0
Gobstopper	0	1	0	0	0	0
Haribo Gold Bears	0	1	0	0	0	0
Haribo Sour Bears	0	1	0	0	0	0
Haribo Twin Snakes	0	1	0	0	0	0
Jawbusters	0	1	0	0	0	0
Laffy Taffy	0	1	0	0	0	0
Lemonhead	0	1	0	0	0	0
Lifesavers big ring gummies	0	1	0	0	0	0
Mike & Ike	0	1	0	0	0	0
Nerds	0	1	0	0	0	0
Nik L Nip	0	1	0	0	0	0
Now & Later	0	1	0	0	0	0
Pop Rocks	0	1	0	0	0	0
Red vines	0	1	0	0	0	0
Ring pop	0	1	0	0	0	0
Runts	0	1	0	0	0	0
Skittles original	0	1	0	0	0	0
Skittles wildberry	0	1	0	0	0	0
Smarties candy	0	1	0	0	0	0
Sour Patch Kids	0	1	0	0	0	0
Sour Patch Tricksters	0	1	0	0	0	0
Starburst	0	1	0	0	0	0
Strawberry bon bons	0	1	0	0	0	0
Super Bubble	0	1	0	0	0	0
Swedish Fish	0	1	0	0	0	0
Tootsie Pop	1	1	0	0	0	0
Trolli Sour Bites	0	1	0	0	0	0
Twizzlers	0	1	0	0	0	0
Warheads	0	1	0	0	0	0
Welch's Fruit Snacks	0	1	0	0	0	0
	crisped rice	wafer	hard bar	pluribus	sugar	percent
Air Heads	0	0	0	0	0	0.906
Caramel Apple Pops	0	0	0	0	0	0.604
Chewey Lemonhead Fruit Mix	0	0	0	1	0	0.732

	0	0	0	1	0.046
Chiclets	0	0	0	1	0.732
Dots	0	0	0	0	0.732
Dum Dums	0	1	0	1	0.127
Fruit Chews	0	0	0	0	0.732
Fun Dip	0	1	0	1	0.906
Gobstopper	0	1	0	1	0.465
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

pricepercent winpercent

Air Heads	0.511	52.34146
Caramel Apple Pops	0.325	34.51768
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 Sour Bears	0.465	51.41243
Haribo Twin Snakes	0.465	42.17877
Jawbusters	0.511	28.12744
Laffy Taffy	0.116	41.38956
Lemonhead	0.104	39.14106
Lifesavers big ring gummies	0.279	52.91139
Mike & Ike	0.325	46.41172
Nerds	0.325	55.35405
Nik L Nip	0.976	22.44534
Now & Later	0.325	39.44680
Pop Rocks	0.837	41.26551
Red vines	0.116	37.34852
Ring pop	0.965	35.29076
Runts	0.279	42.84914
Skittles original	0.220	63.08514
Skittles wildberry	0.220	55.10370
Smarties candy	0.116	45.99583
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
Super Bubble	0.116	27.30386
Swedish Fish	0.755	54.86111
Tootsie Pop	0.325	48.98265
Trolli Sour Bites	0.255	47.17323
Twizzlers	0.116	45.46628
Warheads	0.116	39.01190
Welch's Fruit Snacks	0.313	44.37552

```

fruity.win <- fruity.candy$winpercent
#7. Which mean value is higher
fruity.mean <- mean(fruity.win)
fruity.mean

```

[1] 44.11974

Chocolate is ranked higher than fruity.

Q12. Is this difference statistically significant?

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

Welch Two Sample t-test

```
data: choc.win and fruity.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
```

There is a statistically significant difference.

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

```
library(dplyr)
ord.ind <- order(candy$winpercent)
head(candy[ord.ind,], 5)
```

	chocolate	fruity	caramel	peanut	yalmond	nougat	
Nik L Nip	0	1	0	0	0	0	
Boston Baked Beans	0	0	0	1	0	0	
Chiclets	0	1	0	0	0	0	
Super Bubble	0	1	0	0	0	0	
Jawbusters	0	1	0	0	0	0	
	crispedrice	wafers	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						

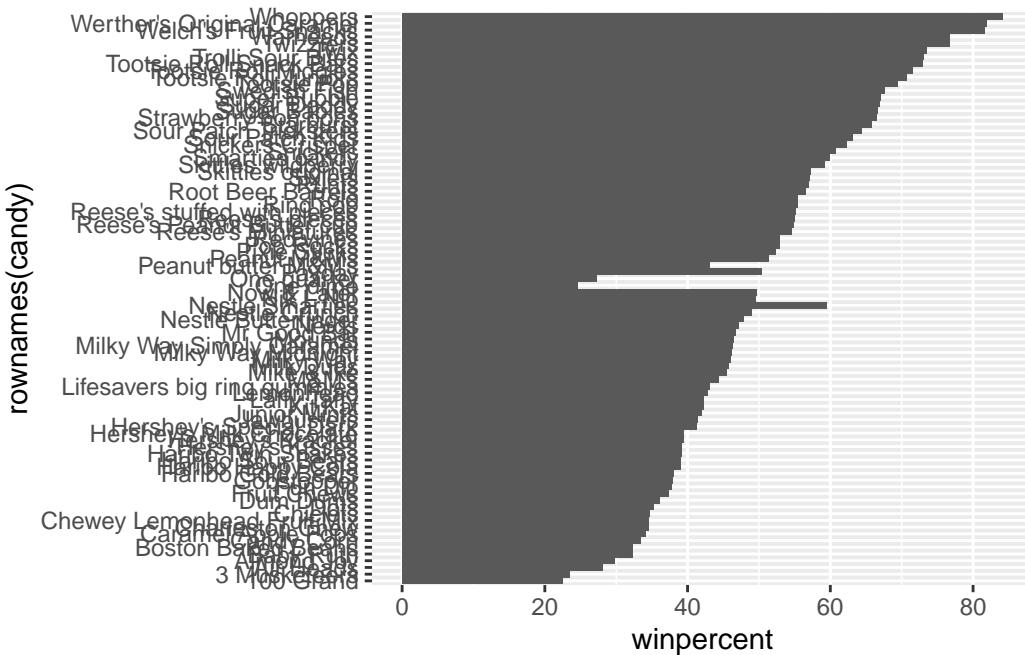
Q14. What are the top 5 all time favorite candy types out of this set?

```
tail(candy[ord.ind,], 5)
```

	chocolate	fruity	caramel	peanut	yalmond	nougat
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
	crispedrice	wafers	hard bar	pluribus	sugar	percent
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	
	price	percent	win	percent		
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				

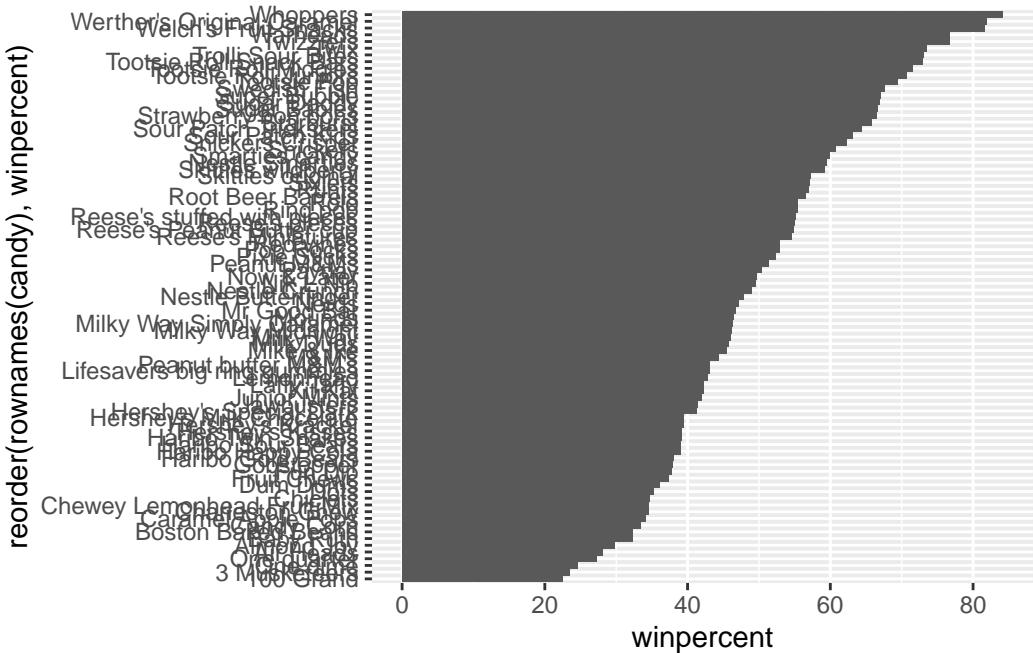
Q15. Make a first barplot of candy ranking based on winpercent values.

```
candy.win <- candy %>% arrange(winpercent)
ggplot(candy.win) +
  aes(winpercent, rownames(candy)) +
  geom_col()
```



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

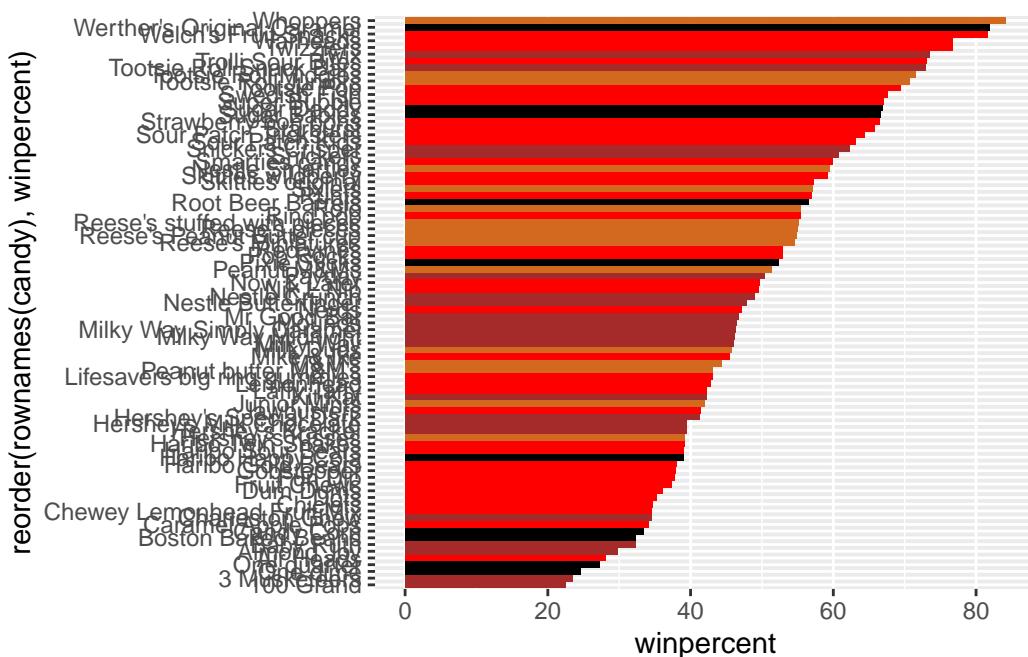
```
candy.win <- candy %>% arrange(winpercent)
ggplot(candy.win) +
  aes(winpercent, reorder(rownames(candy), winpercent)) +
  geom_col()
```



```
my_cols <- rep("black", nrow(candy))
my_cols[candy$chocolate==1] <- "chocolate"
my_cols[candy$bar==1] <- "brown"
my_cols[candy$fruity==1] <- "red"
my_cols
```

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

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



Q17. What is the worst ranked chocolate candy?

Sixlets are the worst ranked chocolate candy

Q18. What is the best ranked fruity candy?

The best ranked fruity candy is starburst.

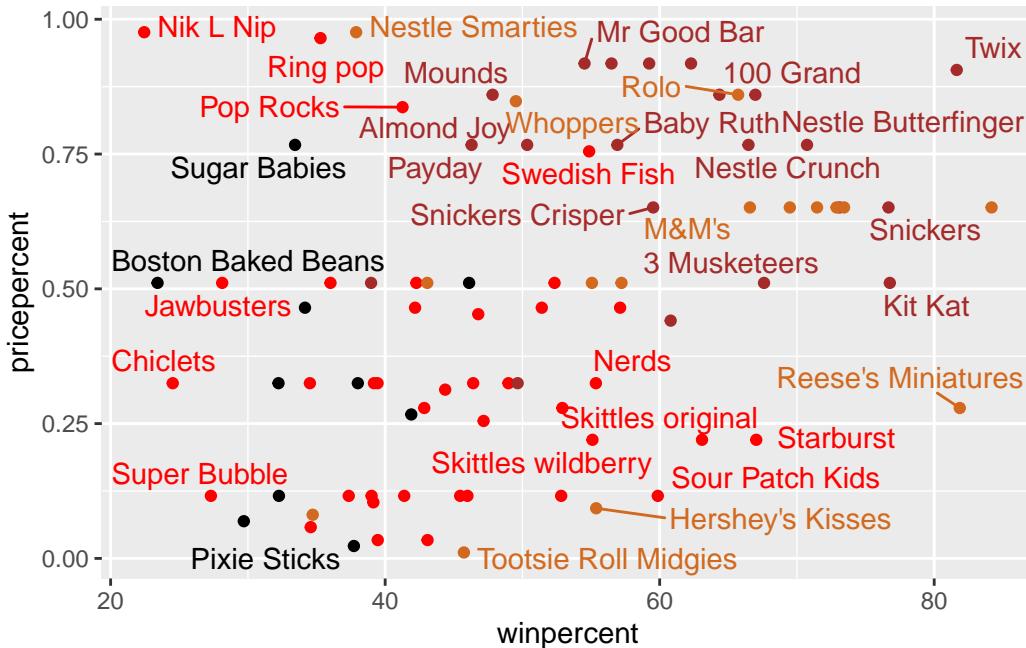
##Winpercent vs Pricepercent

```
library(ggrepel)
```

```
Warning: package 'ggrepel' was built under R version 4.3.3
```

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

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?

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.ind.price <- order(candy$pricepercent)
tail(candy[ord.ind.price,], 5)
```

	chocolate	fruity	caramel	peanut	almond	nougat
Hershey's Special Dark	1	0	0	0	0	0
Mr Good Bar	1	0	0	0	1	0
Ring pop	0	1	0	0	0	0
Nik L Nip	0	1	0	0	0	0
Nestle Smarties	1	0	0	0	0	0
	crispedrice	wafers	hard bar	pluribus	sugarpercent	
Hershey's Special Dark	0	0	1	0	0.430	
Mr Good Bar	0	0	1	0	0.313	

Ring pop	0	1	0	0	0.732
Nik L Nip	0	0	0	1	0.197
Nestle Smarties	0	0	0	1	0.267
	price	percent	win	percent	
Hershey's Special Dark	0.918	59.23612			
Mr Good Bar	0.918	54.52645			
Ring pop	0.965	35.29076			
Nik L Nip	0.976	22.44534			
Nestle Smarties	0.976	37.88719			

##Correlation

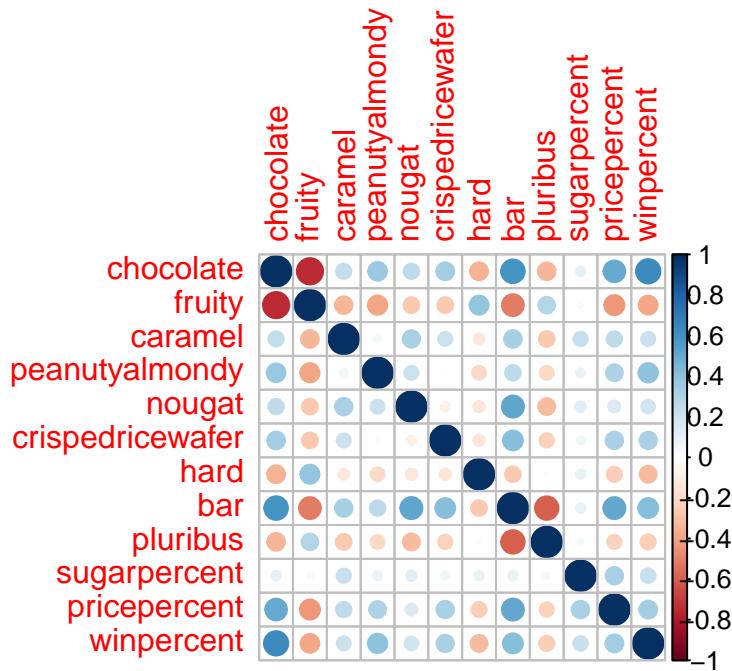
```
cij <- cor(candy)
```

```
library(corrplot)
```

Warning: package 'corrplot' was built under R version 4.3.3

corrplot 0.95 loaded

```
corrplot(cij)
```



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

Fruity and chocolate have strong negative values.

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

Bar and chocolate appear to have a positive correlation.

##Principal Component Analysis

The main function in base R for this is `prcomp()` and we want to set `scale = TRUE` here:

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

Importance of components:

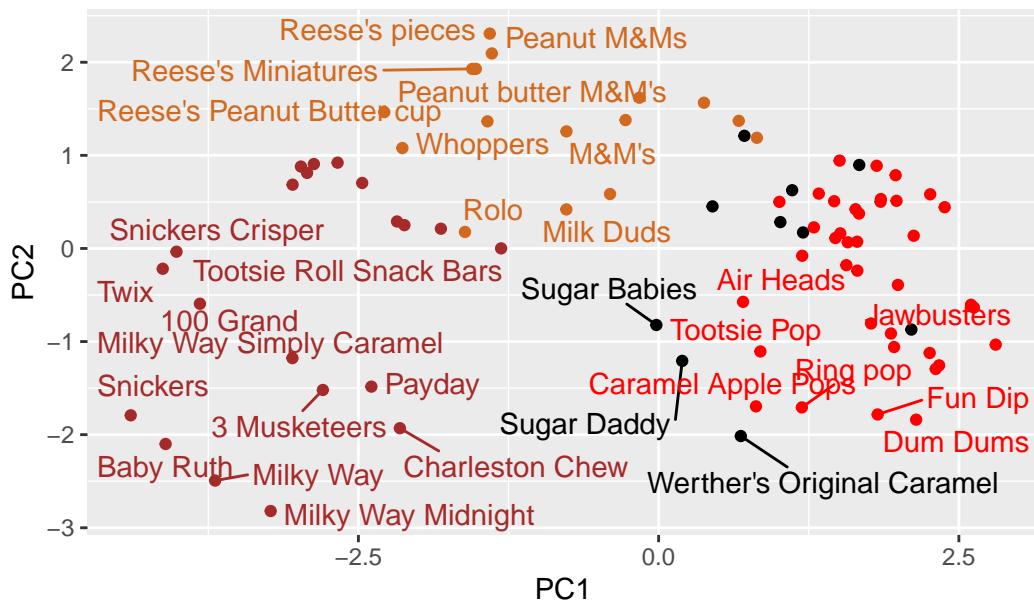
	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	2.0788	1.1378	1.1092	1.07533	0.9518	0.81923	0.81530
Proportion of Variance	0.3601	0.1079	0.1025	0.09636	0.0755	0.05593	0.05539
Cumulative Proportion	0.3601	0.4680	0.5705	0.66688	0.7424	0.79830	0.85369
	PC8	PC9	PC10	PC11	PC12		
Standard deviation	0.74530	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		

Let's look at our first main result figure - the “PC plot” or PC1 vs PC2

```
ggplot(pca$x) +
  aes(PC1, PC2, label = rownames(pca$x)) +
  geom_point(col = my_cols) +
  geom_text_repel(col=my_cols) +
  labs(title = "Ryan's candy analysis")
```

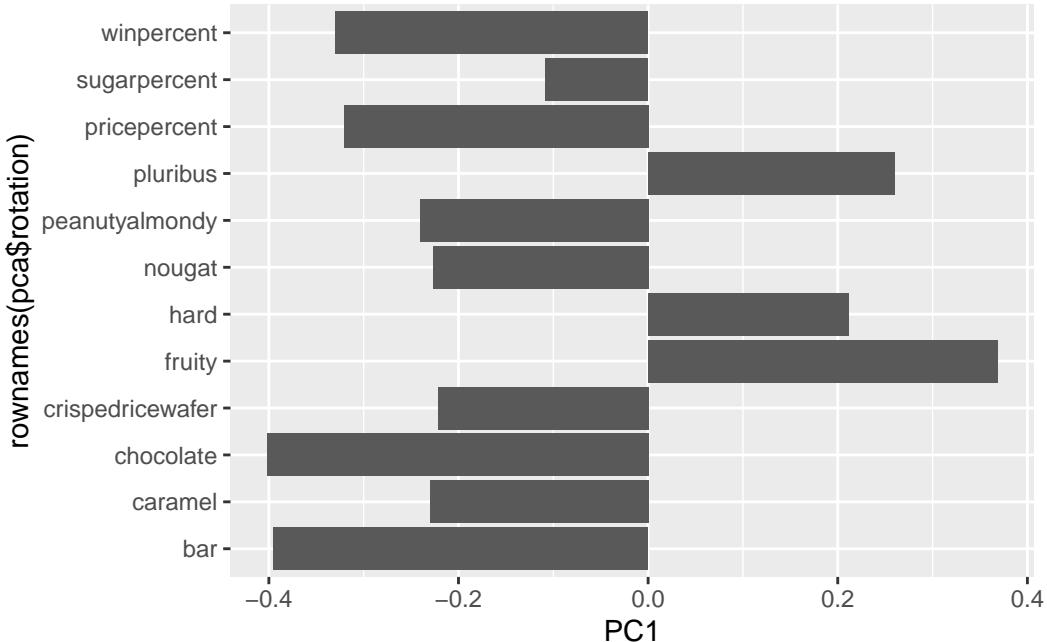
Warning: ggrepel: 54 unlabeled data points (too many overlaps). Consider increasing max.overlaps

## Ryan's candy analysis



Don't forget about your variable "loadings" - how the original variables contribute to your new PCs...

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



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

Pluribus, hard, and fruity are strongly picked up by PC1 in the positive direction. Fruity being detected makes sense to me because it appeared to be a popular choice beforehand, but hard and pluribus was a bit surprising.