

Class 8: Halloween Candy Mini Project

Clarissa Savko (PID: A69028482)

First I have to get my data into R!

```
candy_file <- "candy-data.csv"
candy = read.csv(candy_file, row.names=1)
head(candy)
```

	chocolate	fruity	caramel	peanut	almond	nougat	crisp	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	sugar	percent	price	percent	win	percent
100 Grand	0	1	0		0.732		0.860	66.97	173
3 Musketeers	0	1	0		0.604		0.511	67.60	294
One dime	0	0	0		0.011		0.116	32.26	109
One quarter	0	0	0		0.011		0.511	46.11	650
Air Heads	0	0	0		0.906		0.511	52.34	146
Almond Joy	0	1	0		0.465		0.767	50.34	755

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

```
dim(candy)
```

```
[1] 85 12
```

There are 85 candy types. Q2. How many fruity candy types are in the dataset?

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

```
[1] 38
```

```
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	0.906
Caramel Apple Pops		0	0	0	0.604
Chewey Lemonhead Fruit Mix		0	0	0	1.0732
Chiclets		0	0	0	1.046
Dots		0	0	0	1.0732
Dum Dums		0	1	0	0.732
Fruit Chews		0	0	0	1.0127
Fun Dip		0	1	0	0.732
Gobstopper		0	1	0	1.0906
Haribo Gold Bears		0	0	0	1.0465
Haribo Sour Bears		0	0	0	1.0465
Haribo Twin Snakes		0	0	0	1.0465
Jawbusters		0	1	0	1.0093
Laffy Taffy		0	0	0	0.220
Lemonhead		0	1	0	0.046
Lifesavers big ring gummies		0	0	0	0.267
Mike & Ike		0	0	0	1.0872
Nerds		0	1	0	1.0848
Nik L Nip		0	0	0	1.0197
Now & Later		0	0	0	1.0220
Pop Rocks		0	1	0	1.0604
Red vines		0	0	0	1.0581
Ring pop		0	1	0	0.732
Runts		0	1	0	1.0872
Skittles original		0	0	0	1.0941
Skittles wildberry		0	0	0	1.0941
Smarties candy		0	1	0	0.267
Sour Patch Kids		0	0	0	1.0069
Sour Patch Tricksters		0	0	0	1.0069
Starburst		0	0	0	1.0151
Strawberry bon bons		0	1	0	1.0569
Super Bubble		0	0	0	0.162
Swedish Fish		0	0	0	1.0604
Tootsie Pop		0	1	0	0.604
Trolli Sour Bites		0	0	0	1.0313
Twizzlers		0	0	0	0.220
Warheads		0	1	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			

There are 38 fruity candies in the dataset.

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

```
candy["Reese's Peanut Butter cup", ]$winpercent
```

```
[1] 84.18029
```

My favorite candy is Reeses peanut butter cups! It's `win.percent()` value is 84.18029.

Q4. What is the winpercent value for “Kit Kat”? 76.7686

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

```
[1] 76.7686
```

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

```
[1] 49.6535
```

Q5. What is the winpercent value for “Tootsie Roll Snack Bars”? 49.6535

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

Table 1: Data summary

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

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
chocolate	0	1	0.44	0.50	0.00	0.00	0.00	1.00	1.00	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
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	

Double colon is a shortcut for using a package to avoid using the library function.

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 has a way higher mean and SD than the other columns.

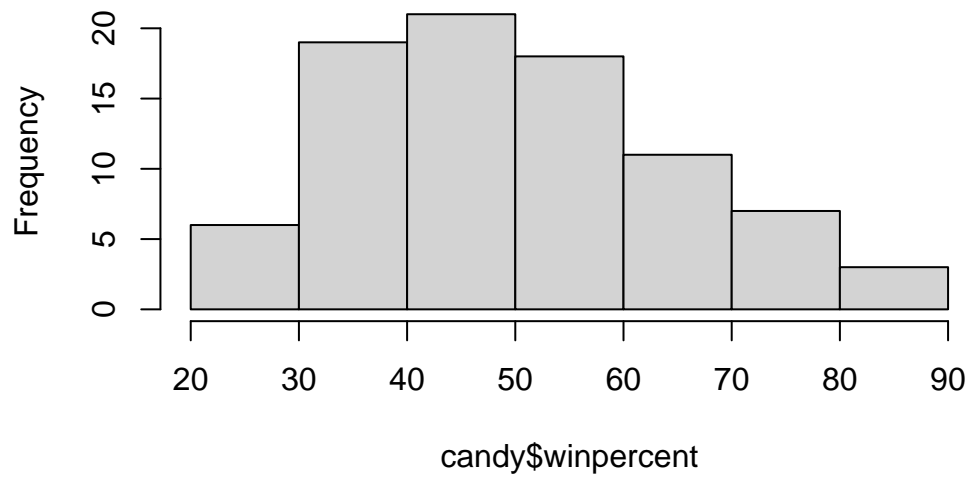
Q7. What do you think a zero and one represent for the candy\$chocolate column?

0 would be if the candy does not contain chocolate (false). 1 would be the candy does contain chocolate (true).

Q8. Plot a histogram of winpercent values

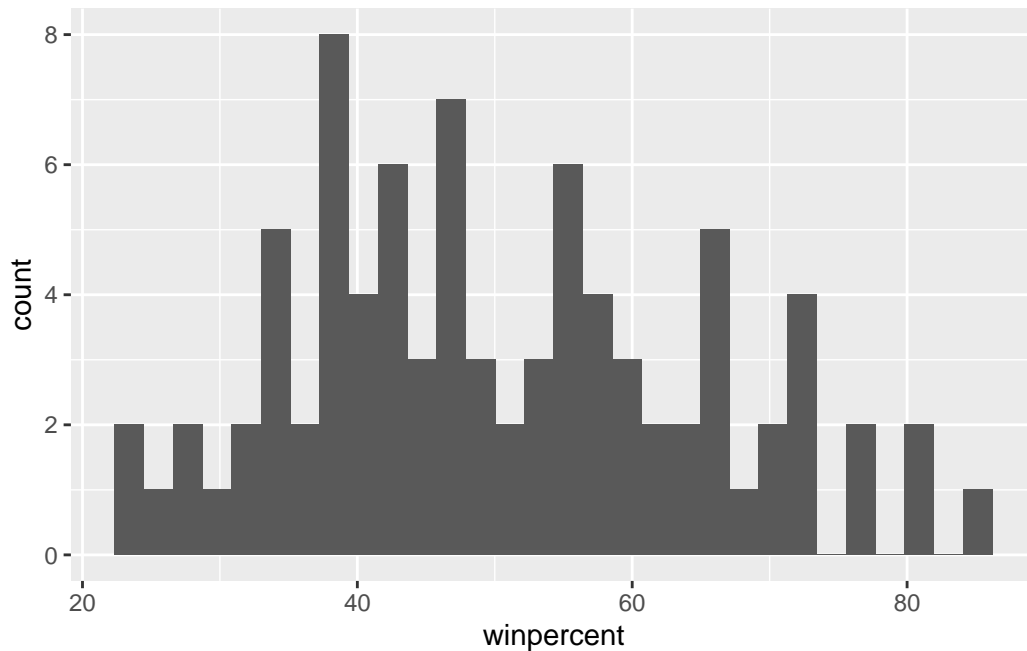
```
hist(candy$winpercent)
```

Histogram of candy\$winpercent



```
library(ggplot2)
ggplot(candy) +
  aes(winpercent)+
  geom_histogram()
```

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



Q9. Is the distribution of winpercent values symmetrical? No, it has a tail to the right.

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

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

```
choc.ind <- as.logical(candy$chocolate)
fruit.ind <- as.logical(candy$fruity)
choc.win <- candy[choc.ind,]$winpercent
fruit.win <- candy[fruit.ind,]$winpercent
mean(choc.win)
```

```
[1] 60.92153
```

```
mean(fruit.win)
```

```
[1] 44.11974
```

choc.win is higher ranked than fruit candy.

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
```

Yes! The p value is 2.871e-08.

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

```
x <- c(5, 2, 3, 6)
sort(x)
```

```
[1] 2 3 5 6
```

```
sort(x, decreasing = T)
```

```
[1] 6 5 3 2
```

```
x
```

```
[1] 5 2 3 6
```

```
order(x)
```

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

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

```
[1] 2 3 5 6
```

```
y <- c("D", "A", "E")
order(y)
```

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

```
y[order(y)]
```

```
[1] "A" "D" "E"
```

```
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
crispedricewafer hard bar pluribus sugarpercent					
Nik L Nip		0	0	0	1 0.197
Boston Baked Beans		0	0	0	1 0.313
Chiclets		0	0	0	1 0.046
Super Bubble		0	0	0	0 0.162
Jawbusters		0	1	0	1 0.093
Root Beer Barrels		0	1	0	1 0.732
Sugar Daddy		0	0	0	0 0.418
One dime		0	0	0	0 0.011
Sugar Babies		0	0	0	1 0.965
Haribo Happy Cola		0	0	0	1 0.465
Caramel Apple Pops		0	0	0	0 0.604
Strawberry bon bons		0	1	0	1 0.569
Sixlets		0	0	0	1 0.220
Ring pop		0	1	0	0 0.732
Chewey Lemonhead Fruit Mix		0	0	0	1 0.732
Red vines		0	0	0	1 0.581
Pixie Sticks		0	0	0	1 0.093
Nestle Smarties		0	0	0	1 0.267
Candy Corn		0	0	0	1 0.906
Charleston Chew		0	0	1	0 0.604
Warheads		0	1	0	0 0.093
Lemonhead		0	1	0	0 0.046
Fun Dip		0	1	0	0 0.732
Now & Later		0	0	0	1 0.220
Dum Dums		0	1	0	0 0.732
Pop Rocks		0	1	0	1 0.604

Laffy Taffy	0	0	0	0	0.220
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[inds,])
```

	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		
Root Beer Barrels	0	0	0		0	0		
	crisp	edrice	wafer	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
Root Beer Barrels		0	1	0		1	0.732	0.069
	winpercent							
Nik L Nip	22.44534							
Boston Baked Beans	23.41782							
Chiclets	24.52499							
Super Bubble	27.30386							
Jawbusters	28.12744							
Root Beer Barrels	29.70369							


```
tail(candy[inds,])
```

	chocolate	fruity	caramel	peanut	almond	nougat
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	rice	wafer	hard	bar	pluribus	sugar	percent
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

	price	percent	win	percent
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		

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

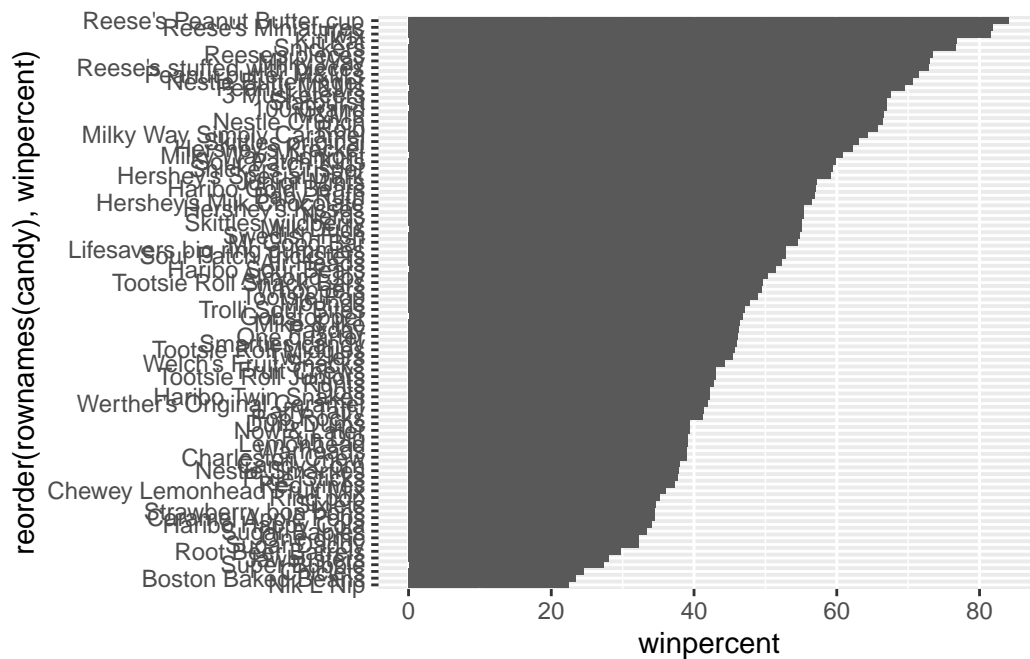
```
candy %>% arrange(winpercent) %>% head(5)
```

	chocolate	fruity	caramel	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
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
	win	percent						
Nik L Nip	22.44	534						
Boston Baked Beans	23.41	782						
Chiclets	24.52	499						
Super Bubble	27.30	386						
Jawbusters	28.12	744						

The 5 least liked candy types are Nik l Nip, Boston Baked Beans, Chiclets, Super Bubble, and Jawbusters. Q14. What are the top 5 all time favorite candy types out of this set? The top 5 candies are Reese's peanut butter cups, Reese's miniatures, Twix, Kit Kat, and Snickers.

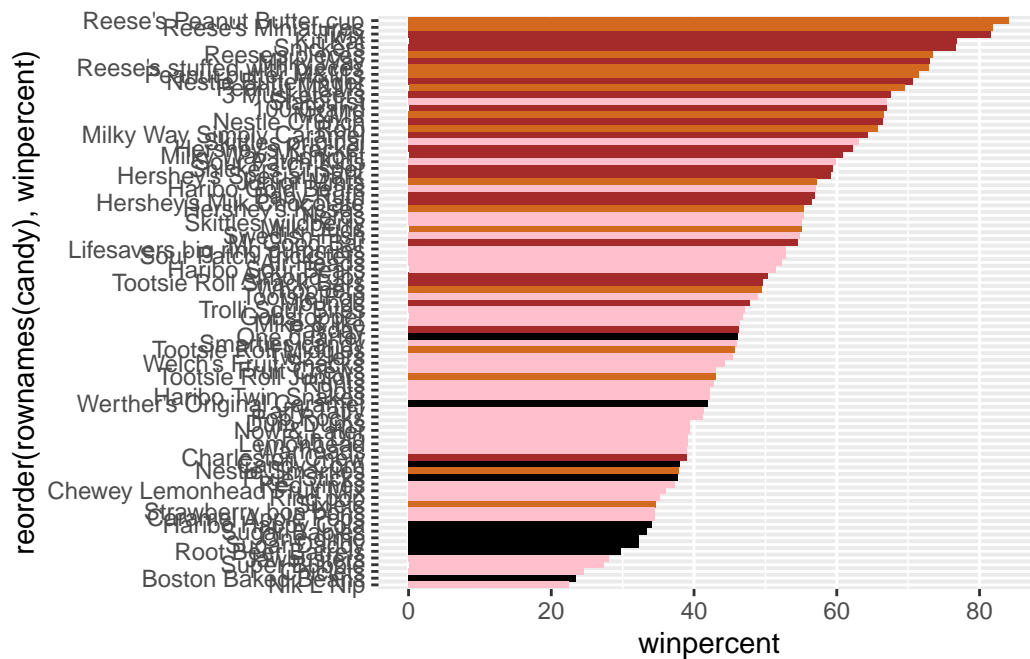
Q15. Make a first barplot of candy ranking based on winpercent values.

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

```
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)
```

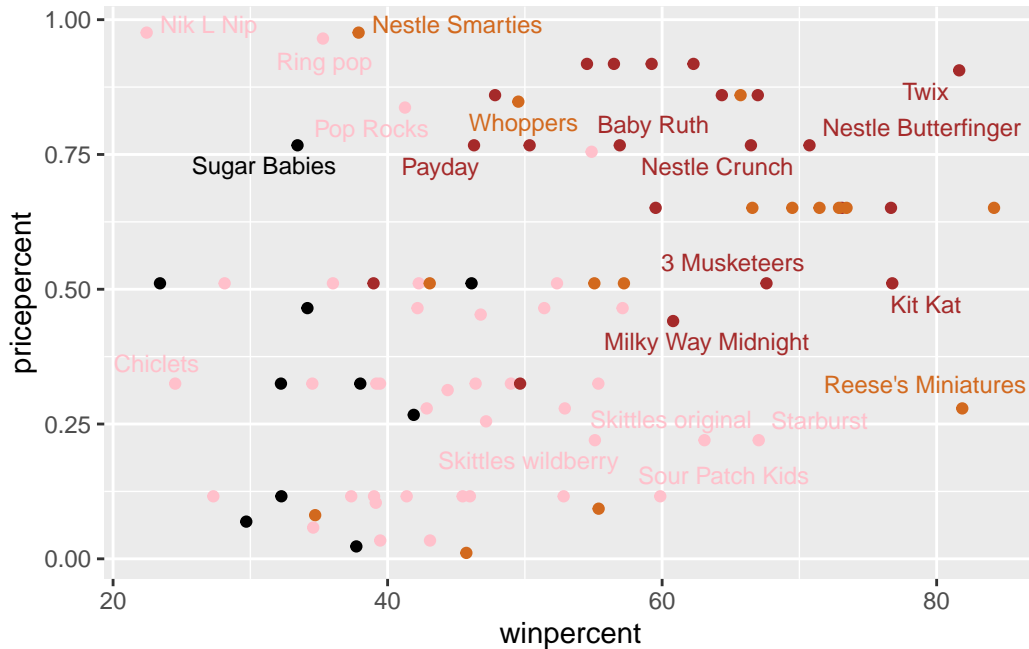


Now, for the first time, using this plot we can answer questions like: - Q17. What is the worst ranked chocolate candy? The worst ranked chocolate candy is Sixlets.

- Q18. What is the best ranked fruity candy? Starburst is the highest ranked fruity candy.

```
library(ggrepel)
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?

Reese's miniature seems to be the highest ranked for the least money.

```
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

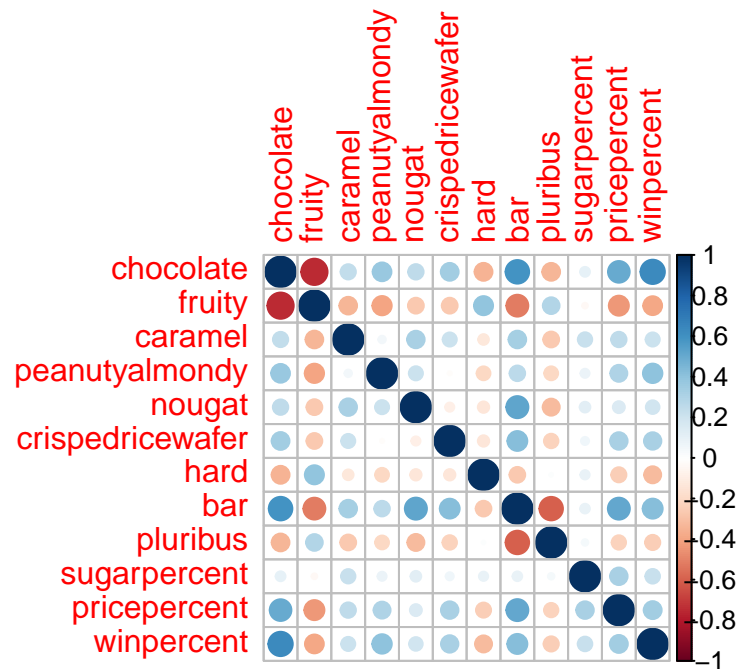
Q20. What are the top 5 most expensive candy types in the dataset and of these which is the least popular?

The top 5 most expensive candies are Nik L Nip, Nestle Smarties, Ring pop, Hershey;s Krackel, and Hershey's milk chocolate. Nik L Nip is the least popular (and most expensive!) with a winpercent of 22.45.

```
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)? Chocolate and fruity are anti-correlated.

Q23. Similarly, what two variables are most positively correlated? Chocolate and winpercent are the most positively correlated.

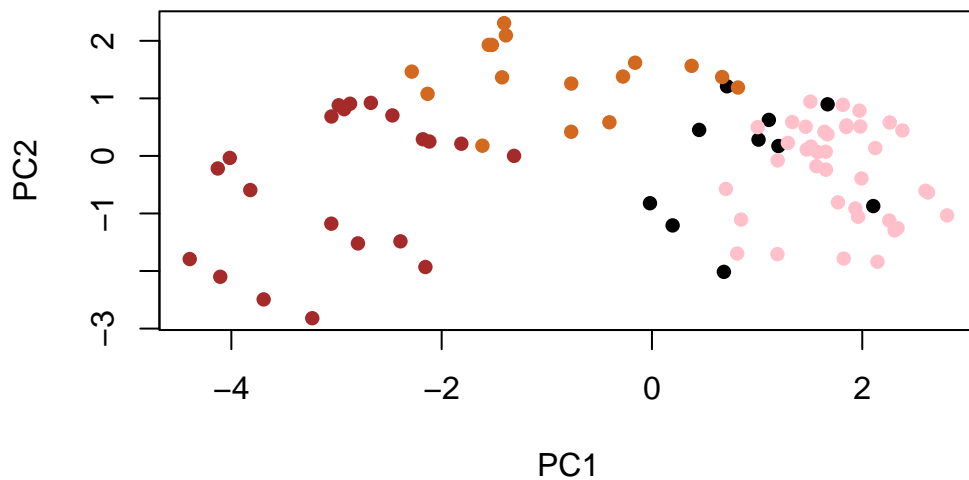
```
pca <- prcomp(candy, scale. = TRUE)
summary(pca)
```

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		

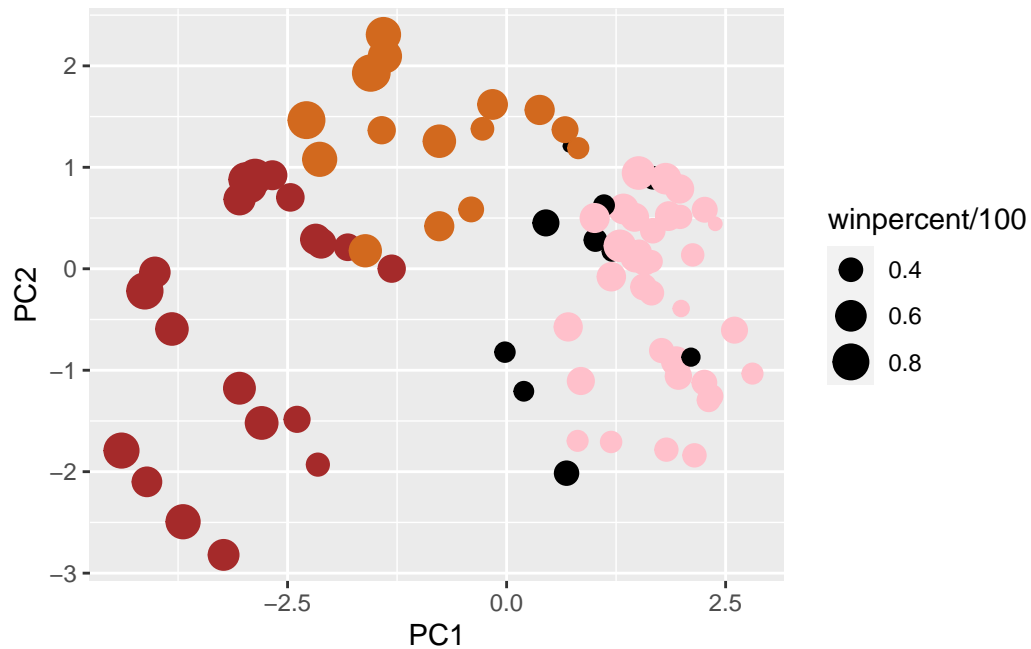
```
plot(pca$x[,1:2], col=my_cols, pch=16)
```



```
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
```

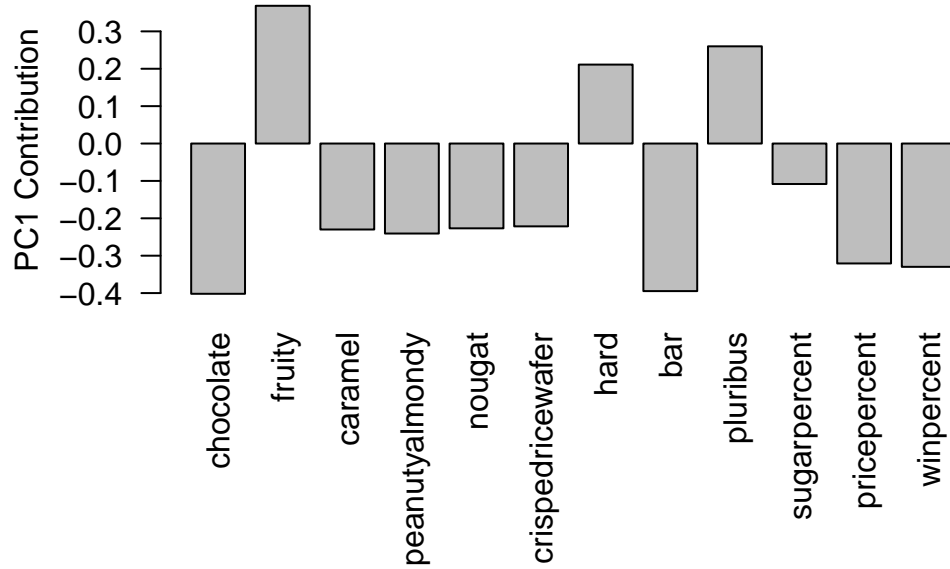



```
library(ggrepel)

#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), chocolate other #(light brown)",
        #caption="Data from 538")

#library(plotly)
#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? Candies that are fruity, hard, and pluribus are picked up strongly in PC1 in the positive direction. This makes sense! Skittles and starbust are hard, fruity, and come with multiple in one package.