Class10

Parnaz Boroon PID:A13557370

```
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
```

There are 85 candy types.

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

```
sum(candy$fruity == 1)
```

```
[1] 38
```

```
colnames(candy)
```

```
 [1] "chocolate"        "fruity"           "caramel"         "peanutyalmondy"
 [5] "nougat"           "crispedricewafer" "hard"            "bar"
 [9] "pluribus"         "sugarpercent"     "pricepercent"    "winpercent"
```

There are 38 fruit candy types.

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

```
[1] 81.64291
```

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

```
candy["Air Heads", "winpercent"]
```

```
[1] 52.34146
```

```
grep("Air", rownames(candy), value = TRUE)
```

```
[1] "Air Heads"
```

Air Heads and 52.34146.

Q4. What is the winpercent value for "Kit Kat"?

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

```
[1] 76.7686
```

```r
grep("Kit Kat", rownames(candy), value = TRUE)
```

```
[1] "Kit Kat"
```

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

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

```
[1] 49.6535
```

```r
grep("Tootsie Roll Snack Bars", rownames(candy), value = TRUE)
```

```
[1] "Tootsie Roll Snack Bars"
```

```r
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
```

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

```
[1] 85
```

```
nrow(candy)
```

```
[1] 85
```

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

```
[1] 50.3168
```

```
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
```

```r
library("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	
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 winpercent row looks to be on a different scale.

Q7. What do you think a zero and one represent for the candy$chocolate column?
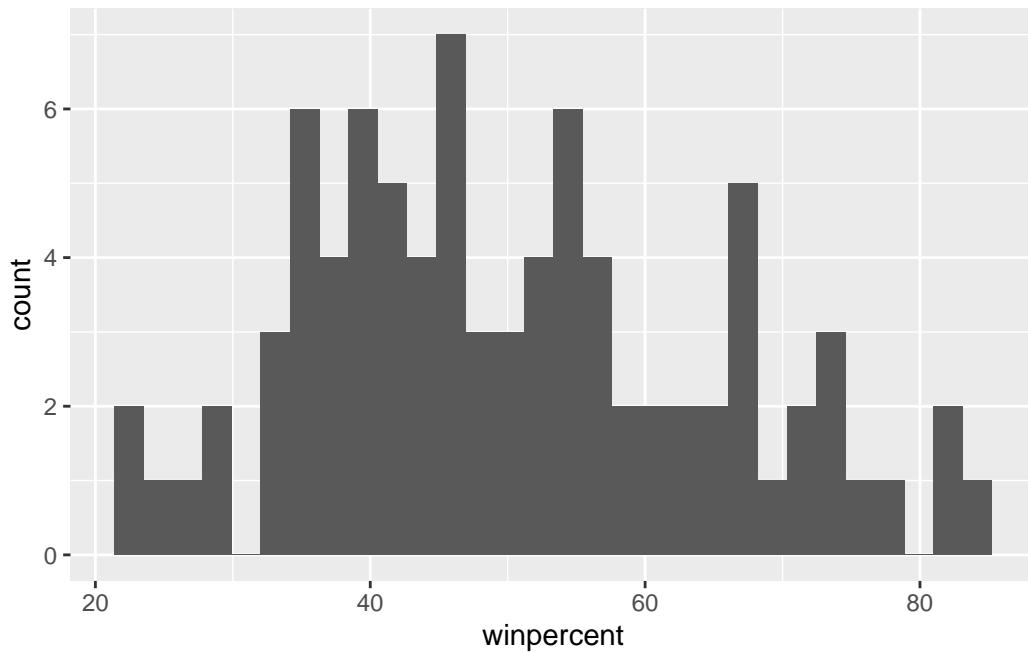
0 and 1 means either candy or chocolate. 1 is chocolate

**Q8. Plot a histogram of winpercent values**
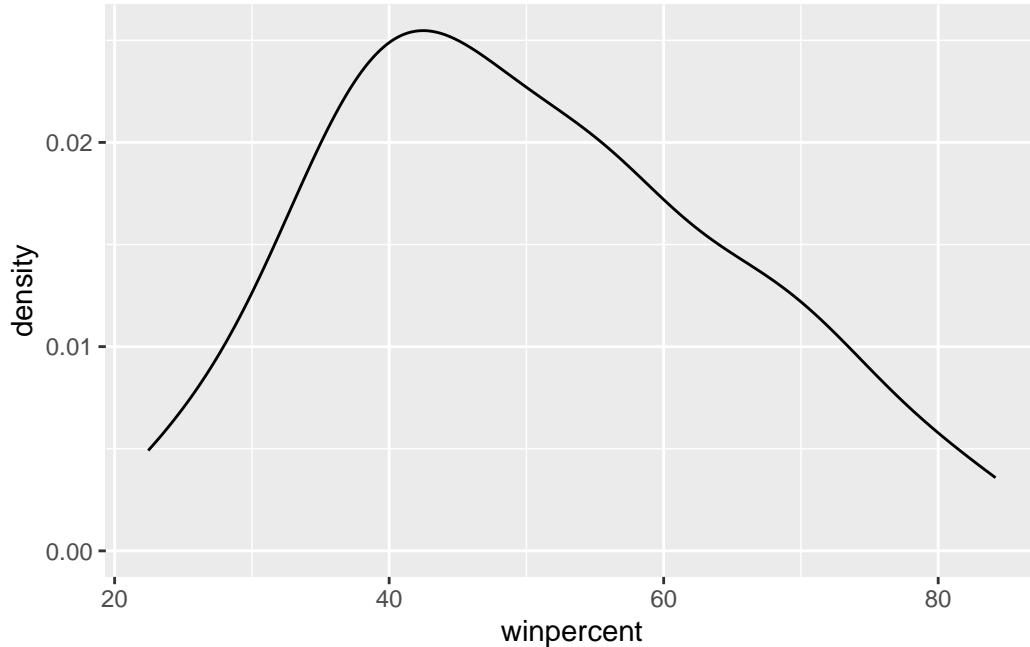
```
library(ggplot2)

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

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

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%?

```
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	peanutyalmondy	nougat
100 Grand	1	0	1	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

	crispedricewafer	hard	bar	pluribus	sugarpercent
100 Grand	1	0	1	0	0.732
3 Musketeers	0	0	1	0	0.604
Almond Joy	0	0	1	0	0.465
Baby Ruth	0	0	1	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

```r
# 2. Extract their `winpercent` values
choc.win <- choc.candy$winpercent
choc.win
```

```
 [1] 66.97173 67.60294 50.34755 56.91455 38.97504 55.37545 62.28448 56.49050
 [9] 59.23612 57.21925 76.76860 71.46505 66.57458 55.06407 73.09956 60.80070
[17] 64.35334 47.82975 54.52645 70.73564 66.47068 69.48379 81.86626 84.18029
[25] 73.43499 72.88790 65.71629 34.72200 37.88719 76.67378 59.52925 48.98265
[33] 43.06890 45.73675 49.65350 81.64291 49.52411
```

```r
# 3. Find the mean of these values
mean(choc.win)
```

```
[1] 60.92153
```

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

	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

	crispedricewafer	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

	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

```
fruits.win <- fruits.candy$winpercent
fruits.win
```

```
[1] 52.34146 34.51768 36.01763 24.52499 42.27208 39.46056 43.08892 39.18550
[9] 46.78335 57.11974 51.41243 42.17877 28.12744 41.38956 39.14106 52.91139
[17] 46.41172 55.35405 22.44534 39.44680 41.26551 37.34852 35.29076 42.84914
[25] 63.08514 55.10370 45.99583 59.86400 52.82595 67.03763 34.57899 27.30386
[33] 54.86111 48.98265 47.17323 45.46628 39.01190 44.37552
```

```
mean(fruits.win)
```

```
[1] 44.11974
```

```
# 7. Which mean value is higher?
#Chocolate candy mean is higher.
```

Q12. Is this difference statistically significant?

yes, significantly different

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

```
	Welch Two Sample t-test

data:  choc.win and fruits.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
```

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

```
x<-c(10,2,5,1)
order(x)
```

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

Q14. What are the top 5 all time favorite candy types out of this set?
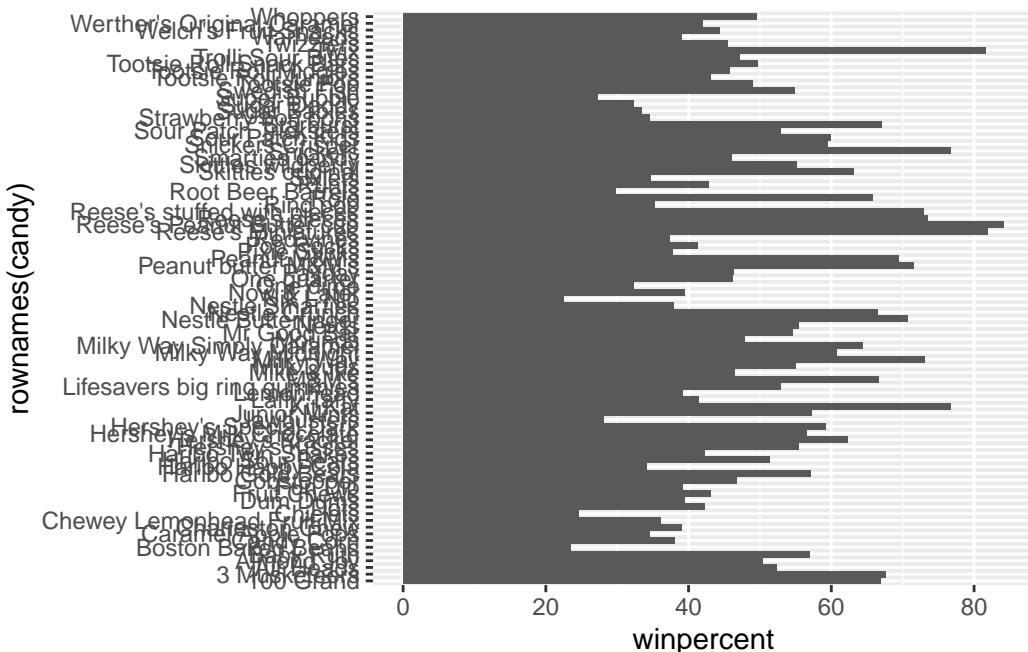
```
ord.ind <- order(candy$winpercent)
head(candy[ord.ind,],5)
```

```
                  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
```

Q15. Make a first barplot of candy ranking based on winpercent values.
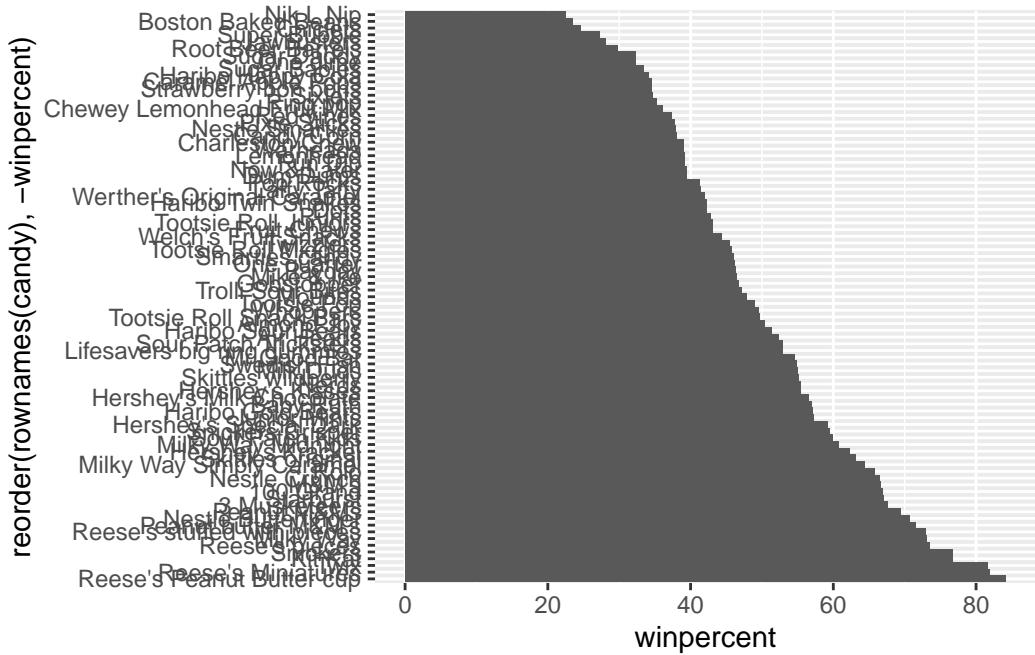
```
library(ggplot2)

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

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

```
library(ggplot2)

ggplot(candy) +
  aes(winpercent, reorder(rownames(candy),-winpercent)) +
  geom_col()
```
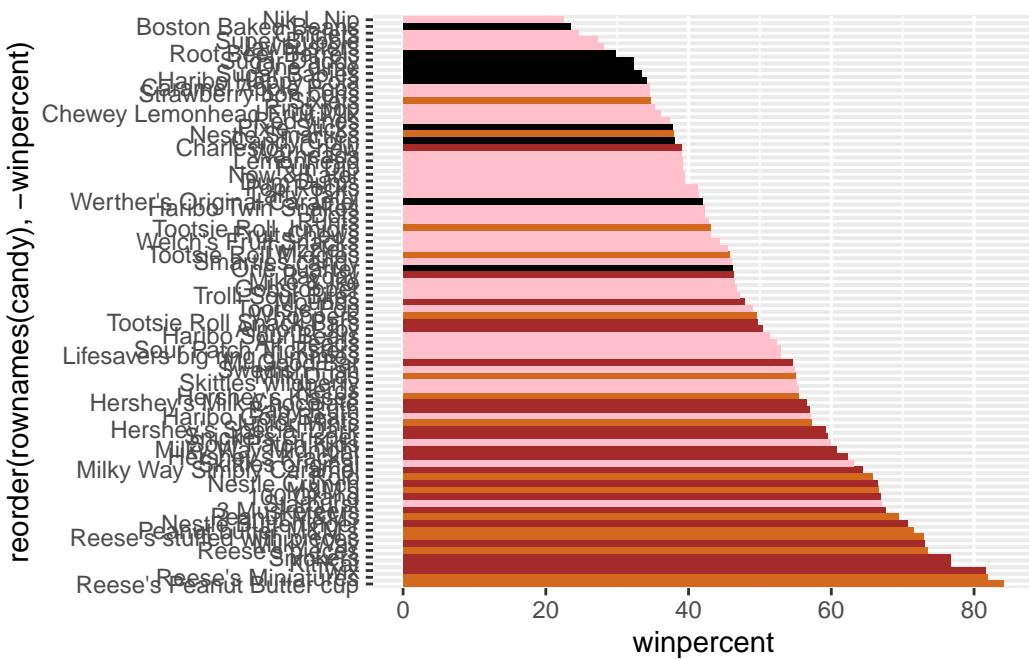
Add some color

```r
my_cols <- rep("black",nrow(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"
my_cols
```

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

```
[85] "chocolate"
```

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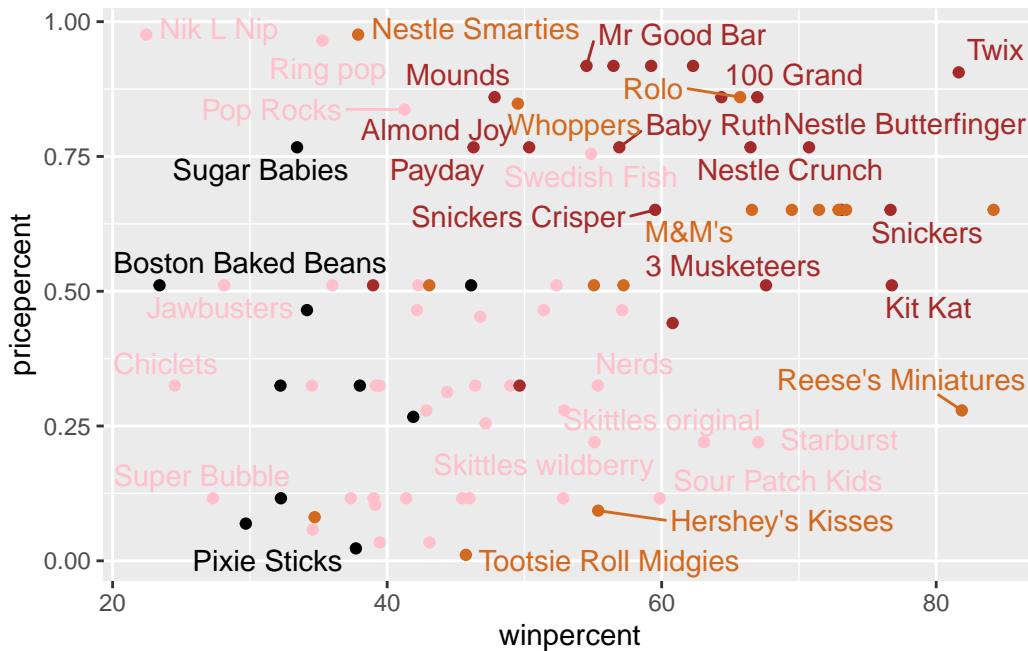

Q17: What is the worst ranked chocolate candy?

Sixlets ## Q18: What is the best ranked fruity candy? Starburst

```r
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)
```

```
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?

```r
ord <- order(candy$pricepercent, decreasing = TRUE)
tail( candy[ord,c(11,12)], n=5 )
```

```
                     pricepercent winpercent
Strawberry bon bons         0.058   34.57899
Dum Dums                    0.034   39.46056
Fruit Chews                 0.034   43.08892
Pixie Sticks                0.023   37.72234
Tootsie Roll Midgies        0.011   45.73675
```

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

```r
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
```

```r
cor(candy)
```

```
                  chocolate       fruity      caramel peanutyalmondy       nougat
chocolate         1.0000000 -0.74172106  0.24987535     0.37782357  0.25489183
fruity           -0.7417211  1.00000000 -0.33548538    -0.39928014 -0.26936712
caramel           0.2498753 -0.33548538  1.00000000     0.05935614  0.32849280
peanutyalmondy    0.3778236 -0.39928014  0.05935614     1.00000000  0.21311310
nougat            0.2548918 -0.26936712  0.32849280     0.21311310  1.00000000
crispedricewafer  0.3412098 -0.26936712  0.21311310    -0.01764631 -0.08974359
hard             -0.3441769  0.39067750 -0.12235513    -0.20555661 -0.13867505
bar               0.5974211 -0.51506558  0.33396002     0.26041960  0.52297636
pluribus         -0.3396752  0.29972522 -0.26958501    -0.20610932 -0.31033884
sugarpercent      0.1041691 -0.03439296  0.22193335     0.08788927  0.12308135
pricepercent      0.5046754 -0.43096853  0.25432709     0.30915323  0.15319643
winpercent        0.6365167 -0.38093814  0.21341630     0.40619220  0.19937530
                  crispedricewafer        hard         bar    pluribus
chocolate               0.34120978 -0.34417691  0.59742114 -0.33967519
fruity                 -0.26936712  0.39067750 -0.51506558  0.29972522
caramel                 0.21311310 -0.12235513  0.33396002 -0.26958501
peanutyalmondy         -0.01764631 -0.20555661  0.26041960 -0.20610932
nougat                 -0.08974359 -0.13867505  0.52297636 -0.31033884
crispedricewafer        1.00000000 -0.13867505  0.42375093 -0.22469338
hard                   -0.13867505  1.00000000 -0.26516504  0.01453172
bar                     0.42375093 -0.26516504  1.00000000 -0.59340892
pluribus               -0.22469338  0.01453172 -0.59340892  1.00000000
sugarpercent            0.06994969  0.09180975  0.09998516  0.04552282
pricepercent            0.32826539 -0.24436534  0.51840654 -0.22079363
winpercent              0.32467965 -0.31038158  0.42992933 -0.24744787
                  sugarpercent pricepercent winpercent
chocolate           0.10416906    0.5046754  0.6365167
fruity             -0.03439296   -0.4309685 -0.3809381
caramel             0.22193335    0.2543271  0.2134163
peanutyalmondy      0.08788927    0.3091532  0.4061922
nougat              0.12308135    0.1531964  0.1993753
crispedricewafer    0.06994969    0.3282654  0.3246797
```

```
hard            0.09180975   -0.2443653  -0.3103816
bar             0.09998516    0.5184065   0.4299293
pluribus        0.04552282   -0.2207936  -0.2474479
sugarpercent    1.00000000    0.3297064   0.2291507
pricepercent    0.32970639    1.0000000   0.3453254
winpercent      0.22915066    0.3453254   1.0000000
```
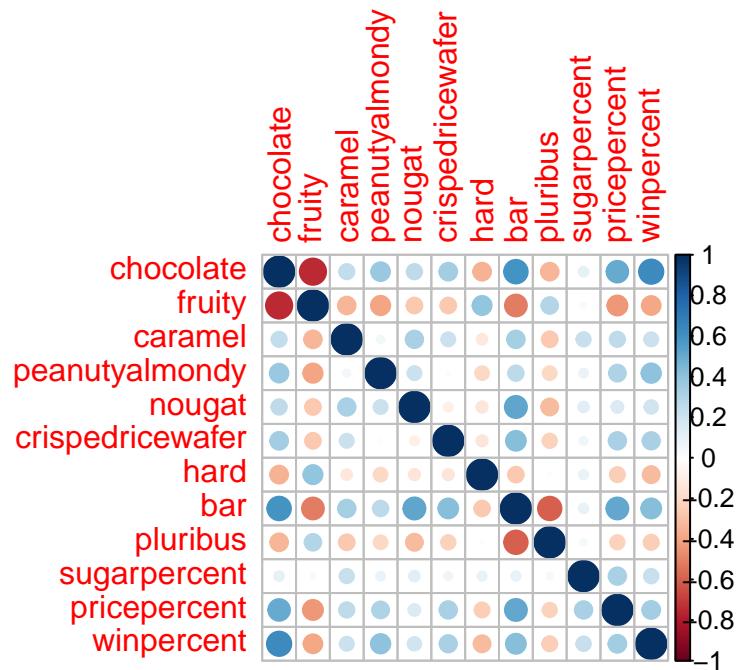
```
cij<- cor(candy)
```

```
library(corrplot)
```

```
corrplot 0.95 loaded
```

```
corrplot(cij)
```


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

Fruity and chocolate

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

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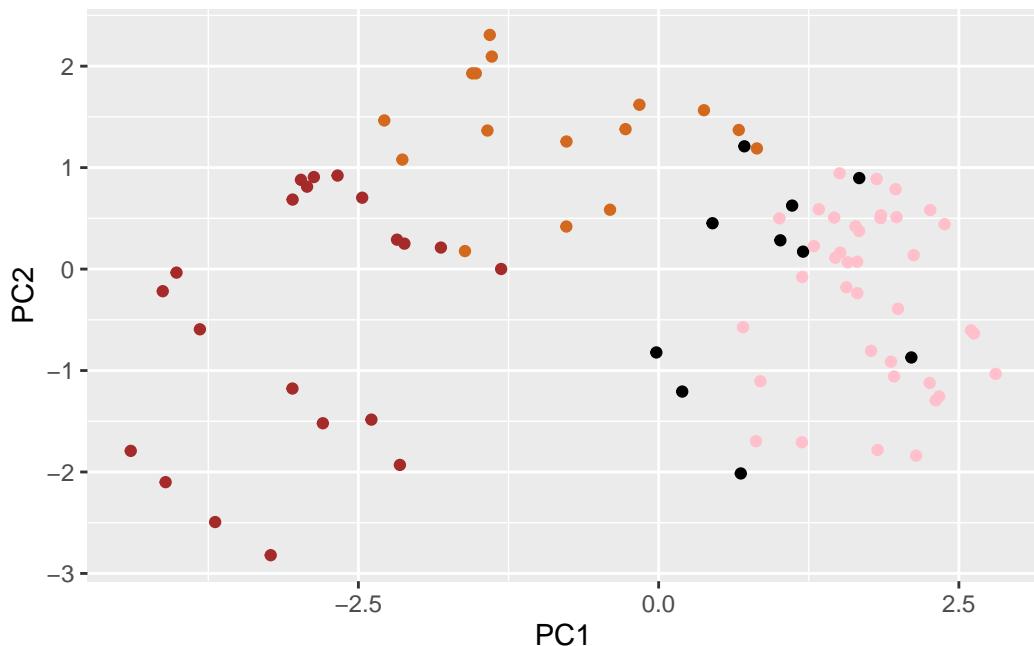
The main function in base R for this `prcomp()`and we want to set `scale=TRUE` here:

```
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
```
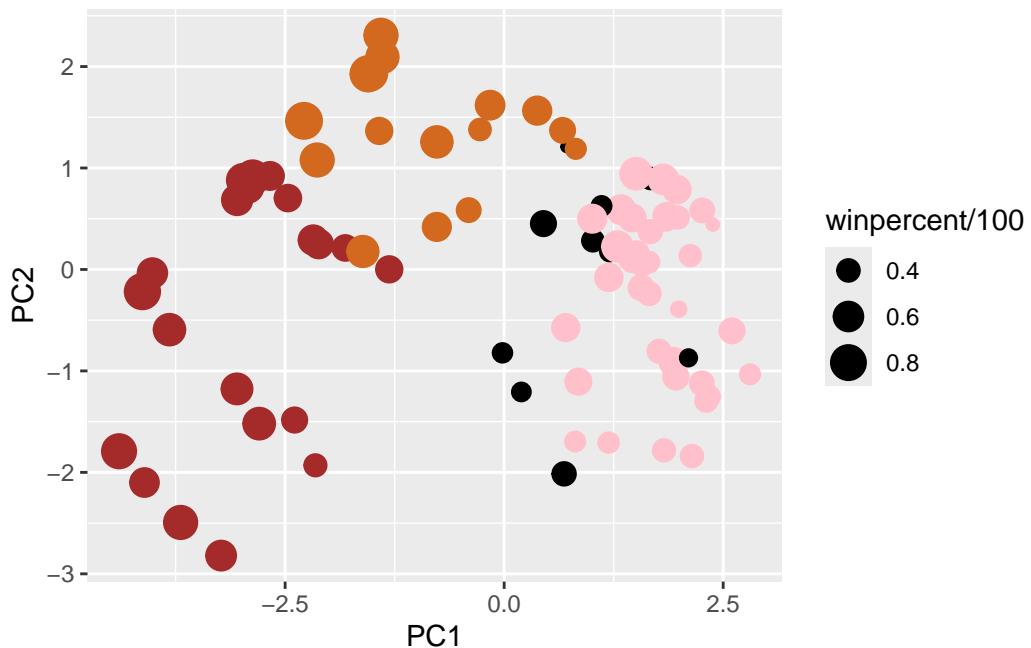
```
ggplot(pca$x)+
  aes(PC1,PC2)+
  geom_point(col=my_cols)
```
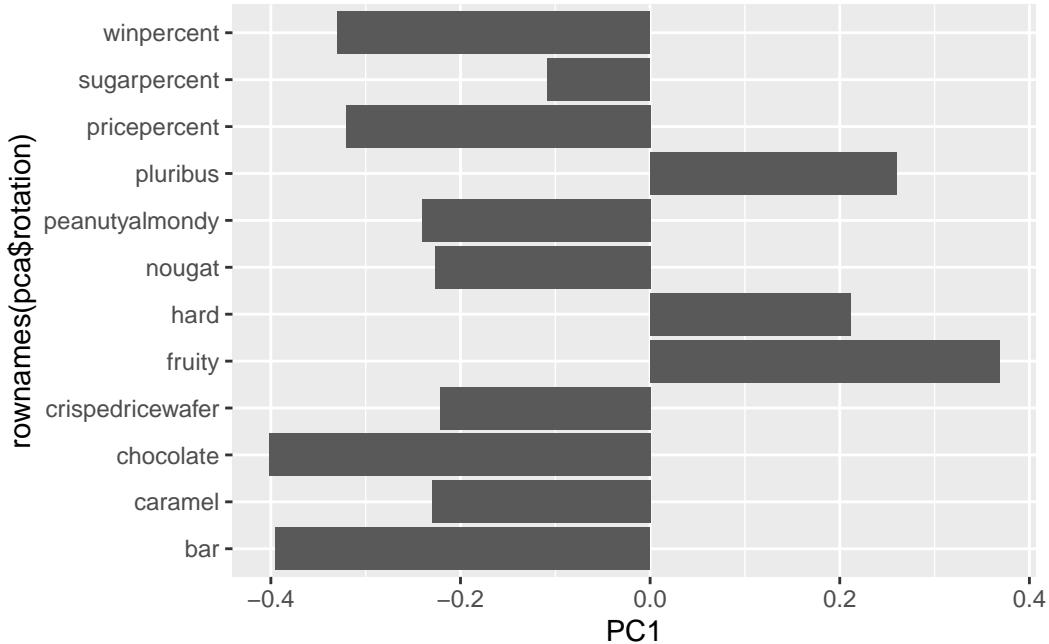
```
# Make a new data-frame with our PCA results and candy data
my_data <- cbind(candy, pca$x[,1:3])

ggplot(my_data) +
        aes(x=PC1, y=PC2,
            size=winpercent/100,
            text=rownames(my_data),
            label=rownames(my_data)) +
        geom_point(col=my_cols)
```



```
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 all in the positive direction which does make sense to me because they all have similar characteristics.