Class 10 Halloween Mini Project

Pamelina Lo (AID: 16735368)

Today is Halloween, an ole Irish holiday, let's celebrate by eating candy.

We will explore some data all about Halloween candy from the 538 website.

1. Importing Candy Data

```r
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?**

Thre are 85 different candy types in this dataset.

```r
nrow(candy)
```

```
[1] 85
```

```r
rownames(candy)
```

```
 [1] "100 Grand"                   "3 Musketeers"
 [3] "One dime"                    "One quarter"
 [5] "Air Heads"                   "Almond Joy"
 [7] "Baby Ruth"                   "Boston Baked Beans"
 [9] "Candy Corn"                  "Caramel Apple Pops"
[11] "Charleston Chew"             "Chewey Lemonhead Fruit Mix"
[13] "Chiclets"                    "Dots"
[15] "Dum Dums"                    "Fruit Chews"
[17] "Fun Dip"                     "Gobstopper"
[19] "Haribo Gold Bears"           "Haribo Happy Cola"
[21] "Haribo Sour Bears"           "Haribo Twin Snakes"
[23] "Hershey's Kisses"            "Hershey's Krackel"
[25] "Hershey's Milk Chocolate"    "Hershey's Special Dark"
[27] "Jawbusters"                  "Junior Mints"
[29] "Kit Kat"                     "Laffy Taffy"
[31] "Lemonhead"                   "Lifesavers big ring gummies"
[33] "Peanut butter M&M's"         "M&M's"
[35] "Mike & Ike"                  "Milk Duds"
[37] "Milky Way"                   "Milky Way Midnight"
[39] "Milky Way Simply Caramel"    "Mounds"
[41] "Mr Good Bar"                 "Nerds"
[43] "Nestle Butterfinger"         "Nestle Crunch"
[45] "Nik L Nip"                   "Now & Later"
[47] "Payday"                      "Peanut M&Ms"
[49] "Pixie Sticks"                "Pop Rocks"
[51] "Red vines"                   "Reese's Miniatures"
[53] "Reese's Peanut Butter cup"   "Reese's pieces"
[55] "Reese's stuffed with pieces" "Ring pop"
[57] "Rolo"                        "Root Beer Barrels"
[59] "Runts"                       "Sixlets"
[61] "Skittles original"           "Skittles wildberry"
[63] "Nestle Smarties"             "Smarties candy"
[65] "Snickers"                    "Snickers Crisper"
[67] "Sour Patch Kids"             "Sour Patch Tricksters"
[69] "Starburst"                   "Strawberry bon bons"
```

```
[71] "Sugar Babies"                 "Sugar Daddy"
[73] "Super Bubble"                 "Swedish Fish"
[75] "Tootsie Pop"                  "Tootsie Roll Juniors"
[77] "Tootsie Roll Midgies"         "Tootsie Roll Snack Bars"
[79] "Trolli Sour Bites"            "Twix"
[81] "Twizzlers"                    "Warheads"
[83] "Welch's Fruit Snacks"         "Werther's Original Caramel"
[85] "Whoppers"
```

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

```r
candy$fruity
```

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

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

```
[1] 38
```

```r
sum(candy$chocolate)
```

```
[1] 37
```

There are 38 fruity candy types in the dataset.

##2. What is your favorite candy?

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

```
[1] 81.64291
```

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

Class Favorite Mentions:

```r
candy["Skittles original","winpercent"]
```

```
[1] 63.08514
```

```r
candy["Skittles original","winpercent"]
```

```
[1] 63.08514
```

My favorite candy:

```r
candy["100 Grand","winpercent"]
```

```
[1] 66.97173
```

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

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

```
[1] 76.7686
```

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

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

```
[1] 49.6535
```

Another way:

```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
```

```
candy |>
  filter(rownames(candy)== "Haribo Happy Cola") |>
  select(winpercent)
```

```
                  winpercent
Haribo Happy Cola   34.15896
```

Class Question: Q. Find furity candy with a winpercent above 50%?

```
candy |>
  filter(winpercent>50) |>
  filter(fruity==1)
```

```
                            chocolate fruity caramel peanutyalmondy nougat
Air Heads                           0      1       0              0      0
Haribo Gold Bears                   0      1       0              0      0
Haribo Sour Bears                   0      1       0              0      0
Lifesavers big ring gummies         0      1       0              0      0
Nerds                               0      1       0              0      0
Skittles original                   0      1       0              0      0
Skittles wildberry                  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
Swedish Fish                        0      1       0              0      0
                            crispedricewafer hard bar pluribus sugarpercent
Air Heads                                  0    0   0        0        0.906
Haribo Gold Bears                          0    0   0        1        0.465
Haribo Sour Bears                          0    0   0        1        0.465
Lifesavers big ring gummies                0    0   0        0        0.267
Nerds                                      0    1   0        1        0.848
Skittles original                          0    0   0        1        0.941
Skittles wildberry                         0    0   0        1        0.941
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
Swedish Fish                               0    0   0        1        0.604
                            pricepercent winpercent
Air Heads                          0.511   52.34146
Haribo Gold Bears                  0.465   57.11974
Haribo Sour Bears                  0.465   51.41243
```

```
Lifesavers big ring gummies        0.279   52.91139
Nerds                              0.325   55.35405
Skittles original                  0.220   63.08514
Skittles wildberry                 0.220   55.10370
Sour Patch Kids                    0.116   59.86400
Sour Patch Tricksters              0.116   52.82595
Starburst                          0.220   67.03763
Swedish Fish                       0.755   54.86111
```

OR this way . . .

```r
top.candy <- candy[candy$winpercent > 50,][candy$fruity==1,]
top.candy[top.candy$fruity == 1,]
```

```
                            chocolate fruity caramel peanutyalmondy nougat
Lifesavers big ring gummies         0      1       0              0      0
Nerds                               0      1       0              0      0
Skittles original                   0      1       0              0      0
Skittles wildberry                  0      1       0              0      0
Sour Patch Kids                     0      1       0              0      0
NA                                 NA     NA      NA             NA     NA
NA.1                               NA     NA      NA             NA     NA
NA.2                               NA     NA      NA             NA     NA
NA.3                               NA     NA      NA             NA     NA
NA.4                               NA     NA      NA             NA     NA
NA.5                               NA     NA      NA             NA     NA
NA.6                               NA     NA      NA             NA     NA
NA.7                               NA     NA      NA             NA     NA
NA.8                               NA     NA      NA             NA     NA
NA.9                               NA     NA      NA             NA     NA
NA.10                              NA     NA      NA             NA     NA
NA.11                              NA     NA      NA             NA     NA
NA.12                              NA     NA      NA             NA     NA
NA.13                              NA     NA      NA             NA     NA
NA.14                              NA     NA      NA             NA     NA
NA.15                              NA     NA      NA             NA     NA
NA.16                              NA     NA      NA             NA     NA
NA.17                              NA     NA      NA             NA     NA
NA.18                              NA     NA      NA             NA     NA
NA.19                              NA     NA      NA             NA     NA
NA.20                              NA     NA      NA             NA     NA
                            crispedricewafer hard bar pluribus sugarpercent
```

```
Lifesavers big ring gummies                0    0   0        0        0.267
Nerds                                      0    1   0        1        0.848
Skittles original                          0    0   0        1        0.941
Skittles wildberry                         0    0   0        1        0.941
Sour Patch Kids                            0    0   0        1        0.069
NA                                        NA   NA  NA       NA           NA
NA.1                                      NA   NA  NA       NA           NA
NA.2                                      NA   NA  NA       NA           NA
NA.3                                      NA   NA  NA       NA           NA
NA.4                                      NA   NA  NA       NA           NA
NA.5                                      NA   NA  NA       NA           NA
NA.6                                      NA   NA  NA       NA           NA
NA.7                                      NA   NA  NA       NA           NA
NA.8                                      NA   NA  NA       NA           NA
NA.9                                      NA   NA  NA       NA           NA
NA.10                                     NA   NA  NA       NA           NA
NA.11                                     NA   NA  NA       NA           NA
NA.12                                     NA   NA  NA       NA           NA
NA.13                                     NA   NA  NA       NA           NA
NA.14                                     NA   NA  NA       NA           NA
NA.15                                     NA   NA  NA       NA           NA
NA.16                                     NA   NA  NA       NA           NA
NA.17                                     NA   NA  NA       NA           NA
NA.18                                     NA   NA  NA       NA           NA
NA.19                                     NA   NA  NA       NA           NA
NA.20                                     NA   NA  NA       NA           NA
                            pricepercent winpercent
Lifesavers big ring gummies        0.279   52.91139
Nerds                              0.325   55.35405
Skittles original                  0.220   63.08514
Skittles wildberry                 0.220   55.10370
Sour Patch Kids                    0.116   59.86400
NA                                    NA         NA
NA.1                                  NA         NA
NA.2                                  NA         NA
NA.3                                  NA         NA
NA.4                                  NA         NA
NA.5                                  NA         NA
NA.6                                  NA         NA
NA.7                                  NA         NA
NA.8                                  NA         NA
NA.9                                  NA         NA
NA.10                                 NA         NA
NA.11                                 NA         NA
NA.12                                 NA         NA
NA.13                                 NA         NA
NA.14                                 NA         NA
NA.15                                 NA         NA
NA.16                                 NA         NA
NA.17                                 NA         NA
NA.18                                 NA         NA
NA.19                                 NA         NA
NA.20                                 NA         NA
```

To get a quick insite into a new dataset some folks like using the skimer package and its `skim()` function

```
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	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
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?**

Yes, it looks like the winpercent variable/column is measures on a different scale to the majority of the other columns in the datasets.

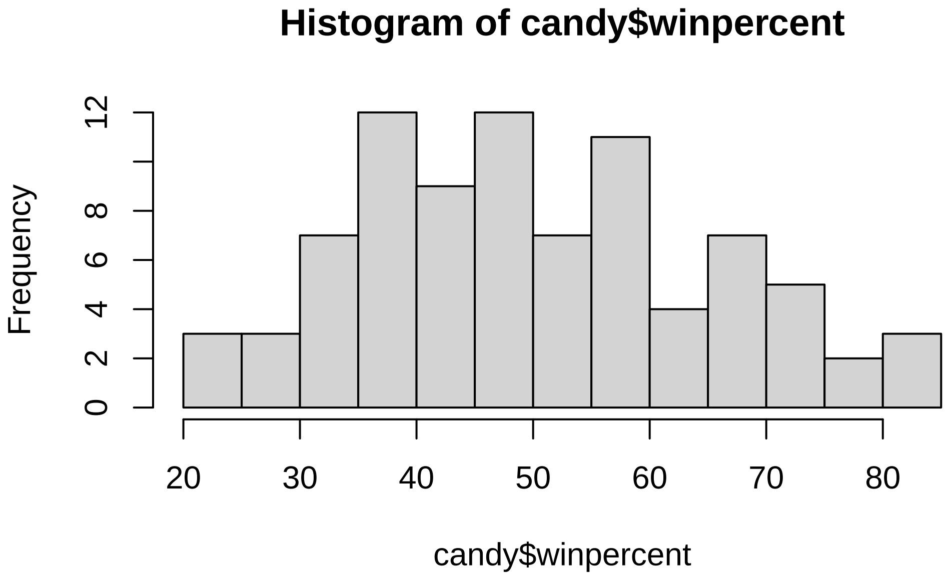
> **Q7. What do you think a zero and one represent for the candy$chocolate column?**

The zeros and one represent True or False statements on if the candy is chocolate or not. If it's classified as chocolate then its a 1 and if its classified as fruity or not chocolate then 0.

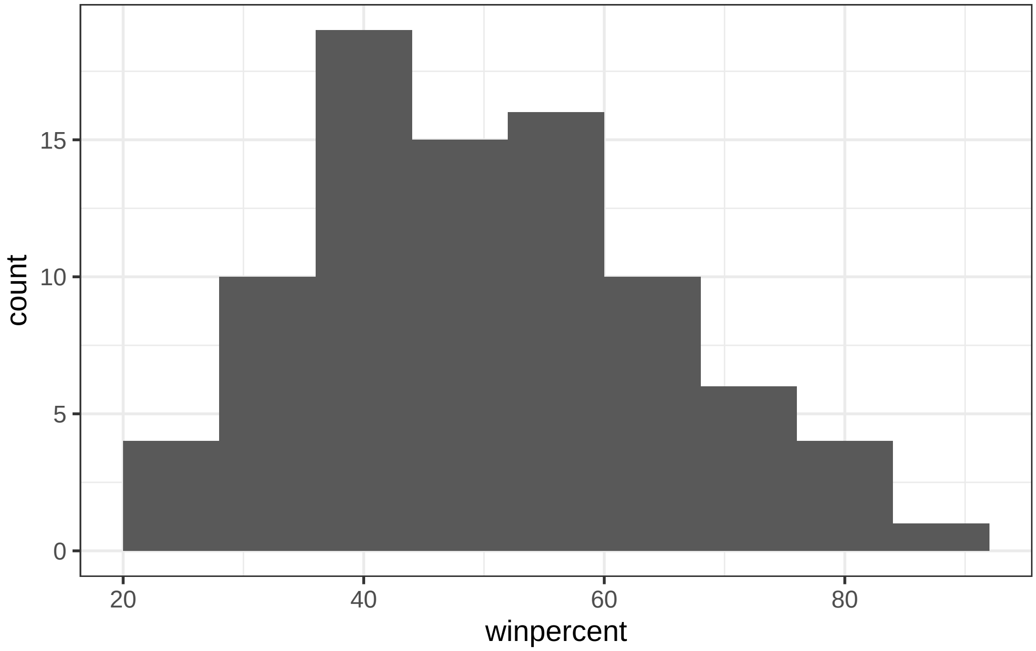
> **Q8. Plot a histogram of winpercent values**

We can do this in a few ways. e.g. the "base" R `hist()` function or with `ggplot()`

```
hist(candy$winpercent, breaks=10)
```



```
library(ggplot2)

ggplot(candy) +
  aes(winpercent) +
  geom_histogram(binwidth = 8) +
  theme_bw()
```



Q9. Is the distribution of winpercent values symmetrical?

No, the distribution of winprecent values is not symmetrical. The distribution appears skewed right.

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

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

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

The center of the distribution is around at 50%, since the mean is 50.32.

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

```
fruit.candy <- candy |>
  filter(fruity==1)

  summary(fruit.candy$winpercent)
```

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

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

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

Chocolate candy appears to be higher ranked than fruit candy.

> **Q12. Is this difference statistically significant?**

```
t.test(candy$chocolate, fruit.candy$pricepercent)
```

```

	Welch Two Sample t-test

data:  candy$chocolate and fruit.candy$pricepercent
t = 1.5336, df = 120.1, p-value = 0.1278
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.02984381  0.23495837
sample estimates:
mean of x mean of y 
0.4352941 0.3327368 
```

No, this difference is not significantly different because the p-value of this t-test is not below 0.05% to be significant.

##3. Overall Candy Rankings >**Q13. What are the five least liked candy types in this set?**

Use sort and order function:

```
play <- c("d","a","c")
sort(play)
```

```
[1] "a" "c" "d"
```

```
order(play)
```

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

```
play[order(play)]
```

```
[1] "a" "c" "d"
```

```
head(candy[order(candy$winpercent),], 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
```

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

```
sort(c(2,5,10), decreasing = T)
```

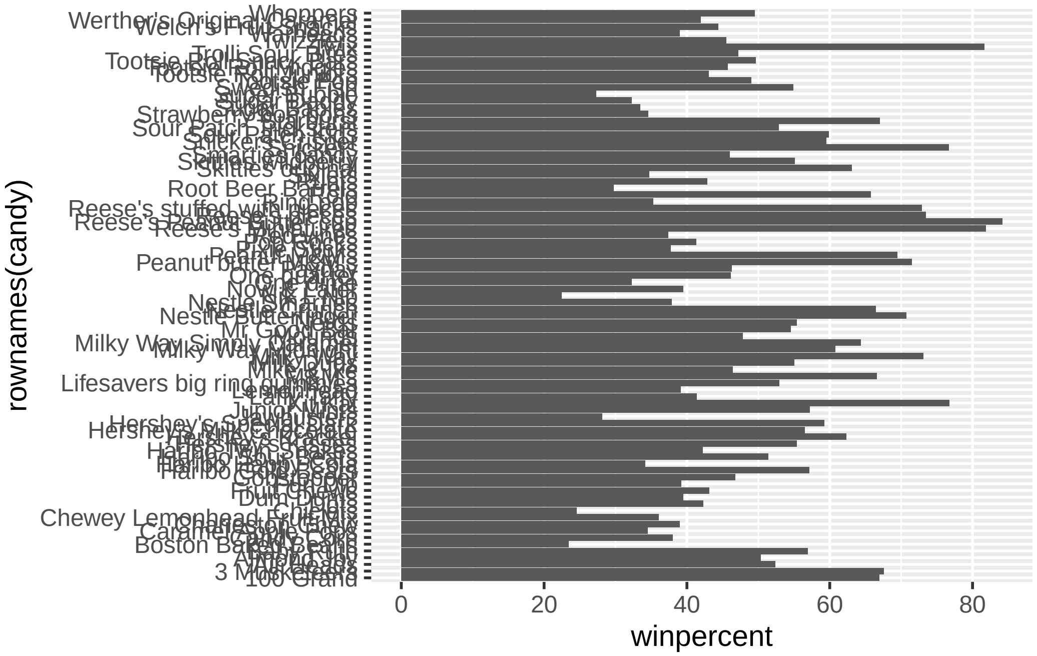
```
[1] 10  5  2
```

```
tail(candy[order(candy$winpercent),], 5)
```

```
                          chocolate fruity caramel peanutyalmondy 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
                          crispedricewafer hard bar pluribus sugarpercent
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
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
```

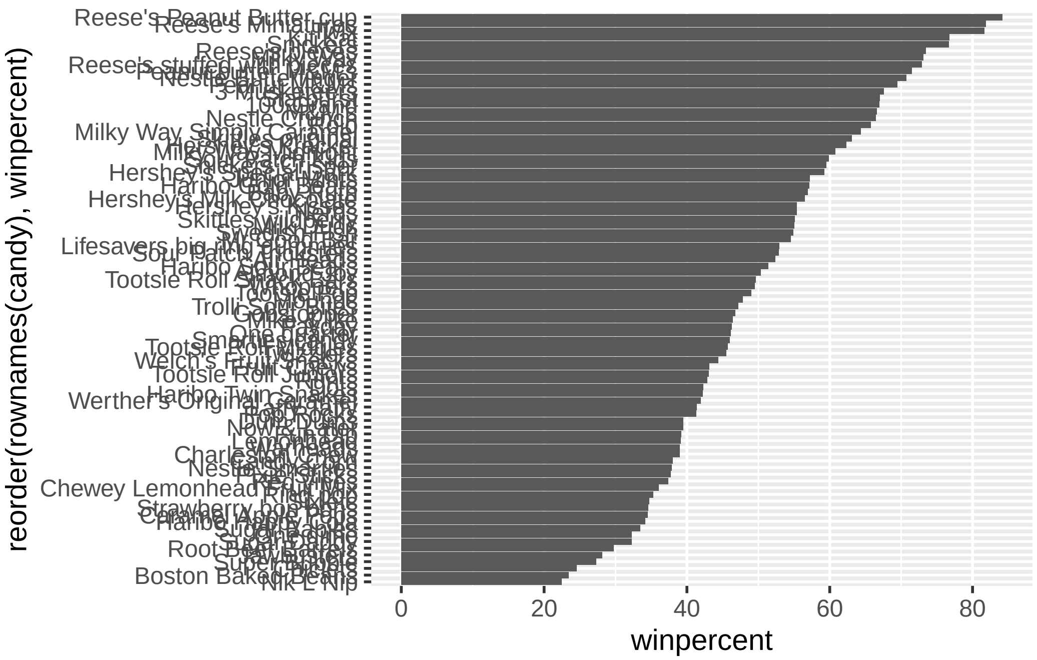
Q15. Make a first barplot of candy ranking based on winpercent values.

```
library(ggplot2)

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



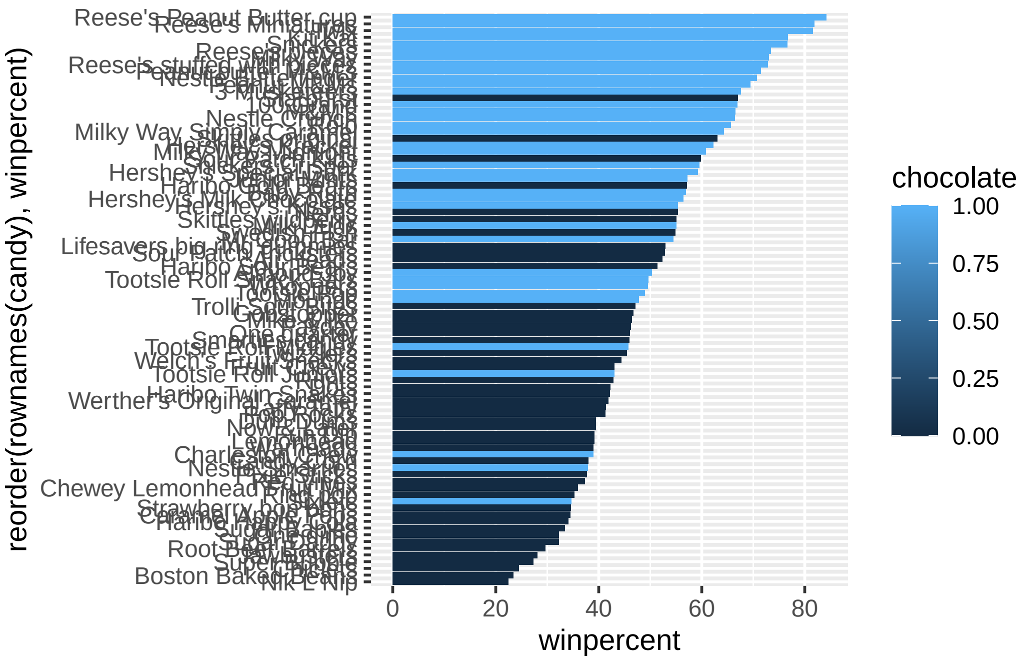
Q16. This is quite ugly, use the reorder() function to get the bars sorted by winpercent?

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



ADD color

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



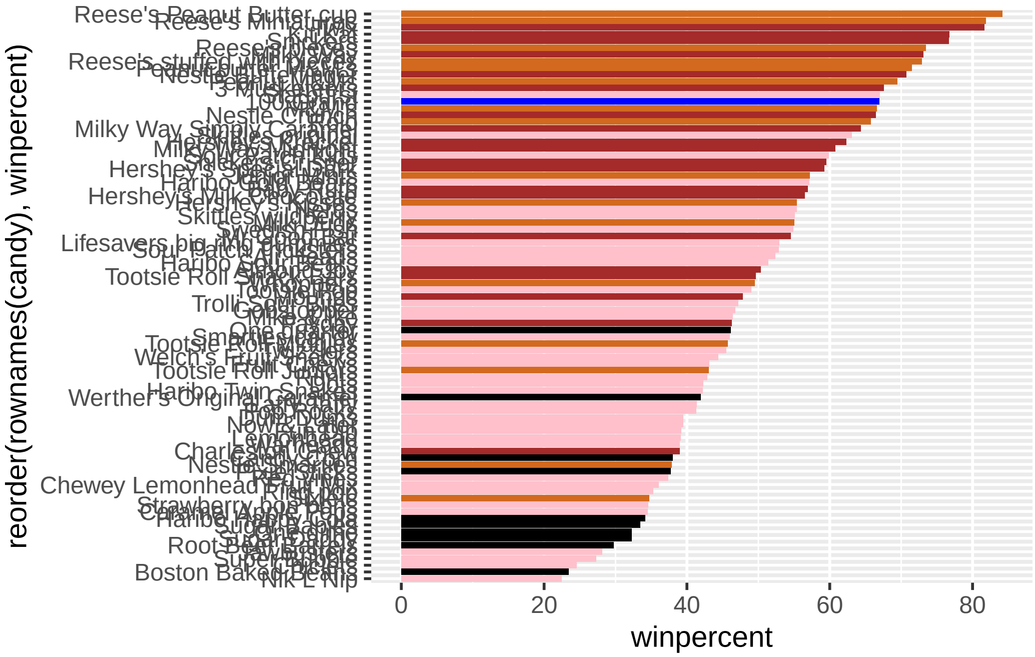
But. . . I want more custom color scheme where I can see both chocolate and bar and fruity etc. all from the one plot. To do this we can roll our own color vector

```
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"
```

or

```
#Place holder color vector:
mycols <- rep("black", nrow(candy))
mycols[as.logical(candy$chocolate)] <- "chocolate"
mycols[as.logical(candy$bar)] <- "brown"
mycols[as.logical(candy$fruity)] <- "pink"

#Use blue for your favorite candy:
mycols[rownames(candy)=="100 Grand"] <- "blue"
```

```
ggplot(candy) +
  aes(x=winpercent,
      y=reorder(rownames(candy),winpercent),
      fill=chocolate) +
  geom_col(fill=mycols)
```



Q17. What is the worst ranked chocolate candy?

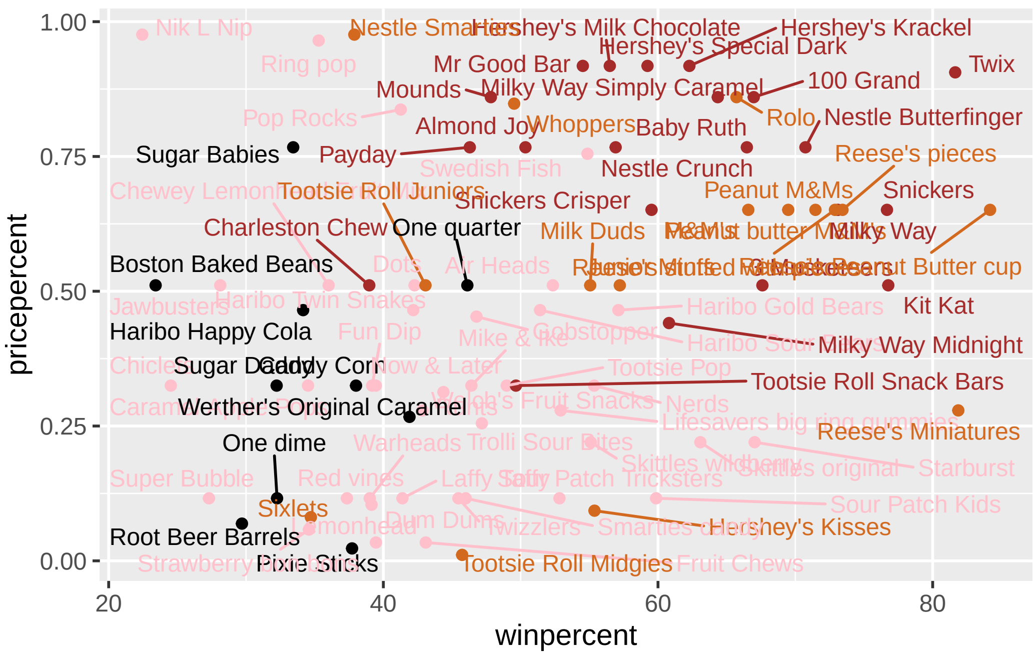
The worst ranked candy is Sixlets.

Q18. What is the best ranked fruity candy?

The best ranked candy is Starburst.

##4. Taking a look at pricepercent

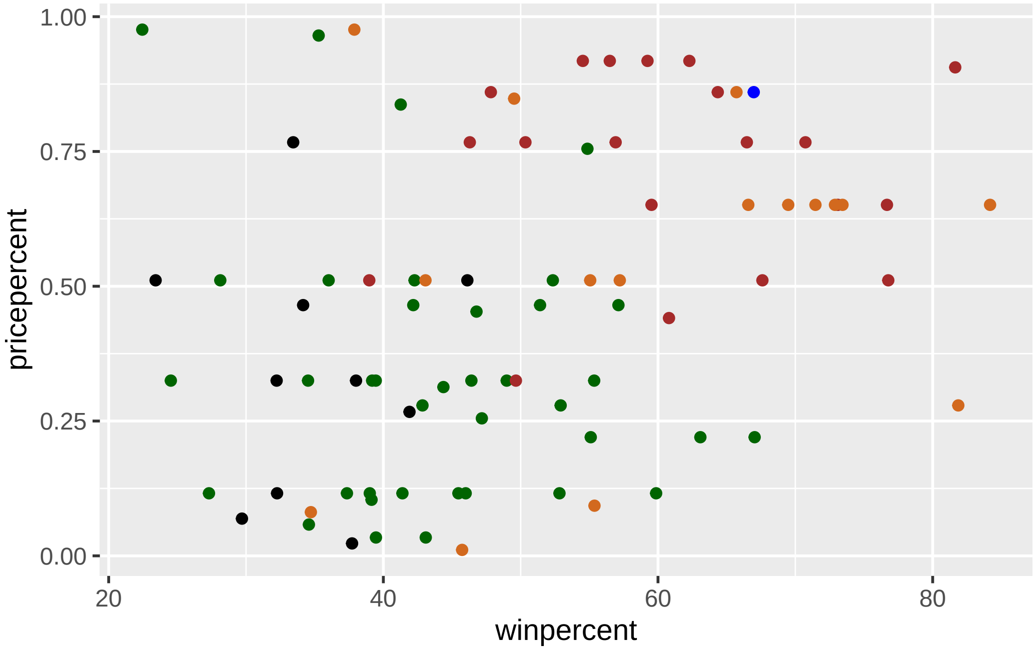
```
library(ggrepel)

# 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 = 29)
```



With the class: Plot of winpercent vs pricepercent to see what would be the best candy to buy

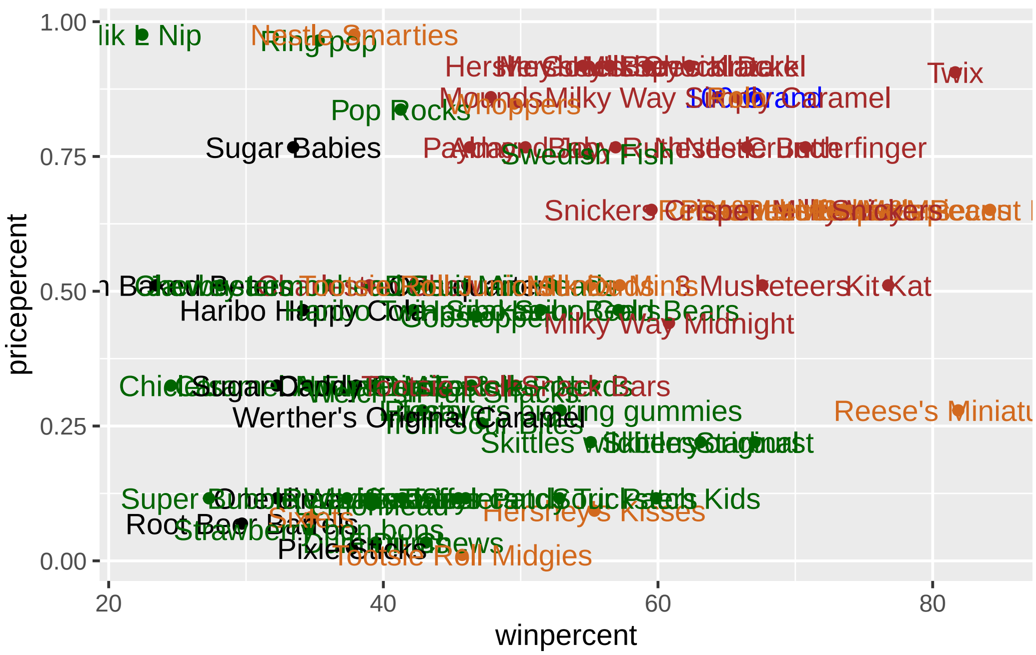
```
mycols[as.logical(candy$fruity)] <- "darkgreen"
```

```
ggplot(candy)+
  aes(winpercent, pricepercent) +
  geom_point(col=mycols)
```



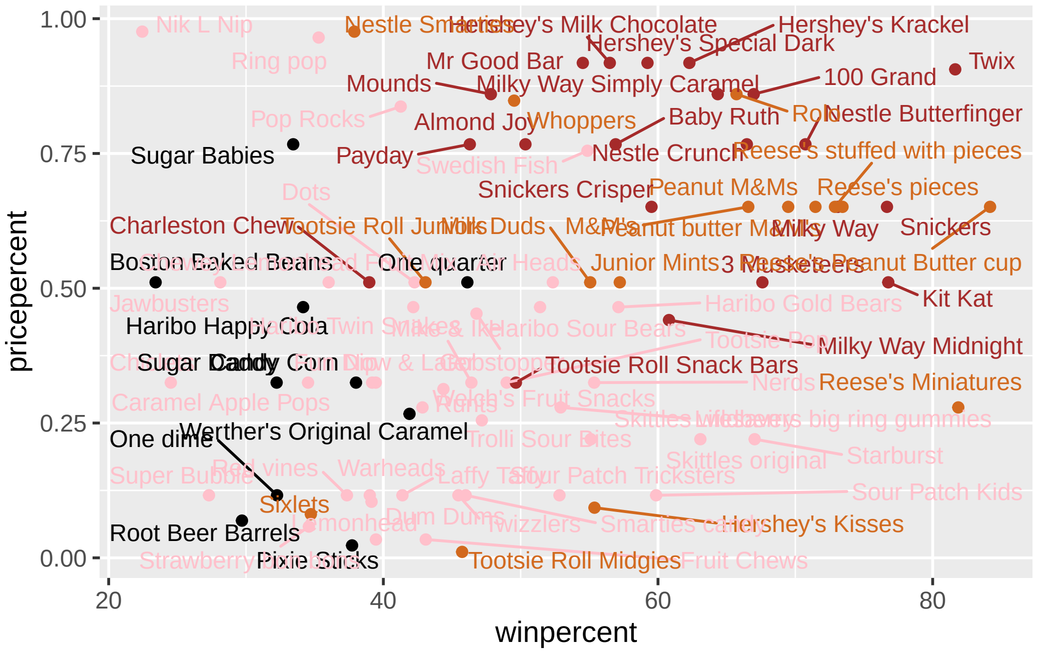
Add labels

```
ggplot(candy)+
  aes(winpercent, pricepercent, label = rownames(candy)) +
  geom_point(col=mycols) +
  geom_text(col=mycols)
```



Make the labels non-overlapping

```r
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 = 29)
```



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?

By looking at the graph, chocolate is the highest candy type in terms for the least money. The candy with the best winpercent for the least amount of money would be Reeses minitures.

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

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

```
                         pricepercent winpercent
Nik L Nip                       0.976   22.44534
Nestle Smarties                 0.976   37.88719
Ring pop                        0.965   35.29076
Hershey's Krackel               0.918   62.28448
Hershey's Milk Chocolate        0.918   56.49050
```

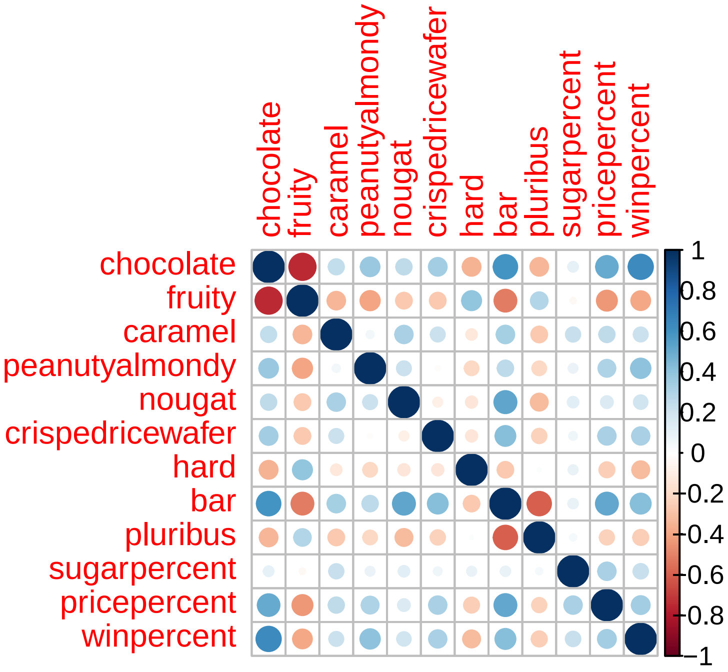
Q21. Make a barplot again with geom_col() this time using pricepercent and then improve this step by step, first ordering the x-axis by value and finally making a so called "dot chat" or "lollipop" chart by swapping geom_col() for geom_point() + geom_segment().

##5. Exploring the correlation structure

```
library(corrplot)
```

```
corrplot 0.95 loaded
```

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



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

Fruit and chocolate are anti-correlated.

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

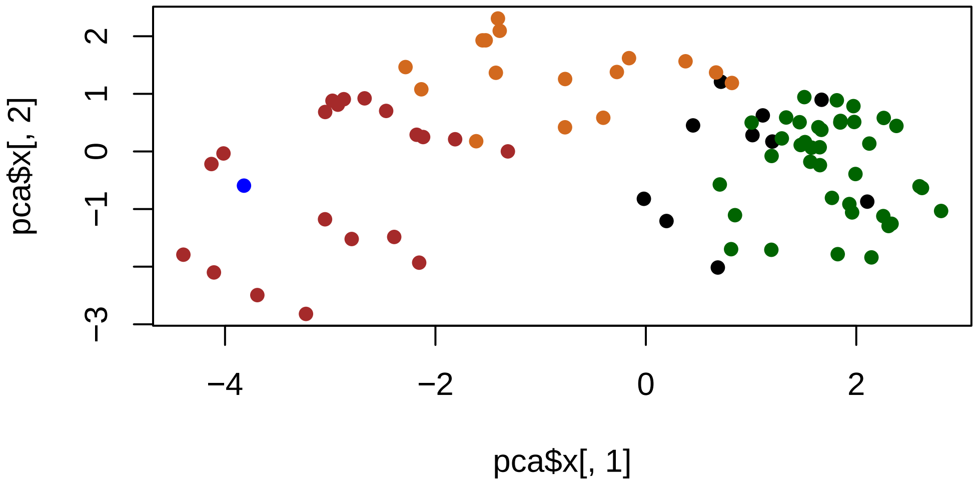
Bar chocolate and chocolate are most positively correlated.

##6. Principal Component Analysis (PCA)

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

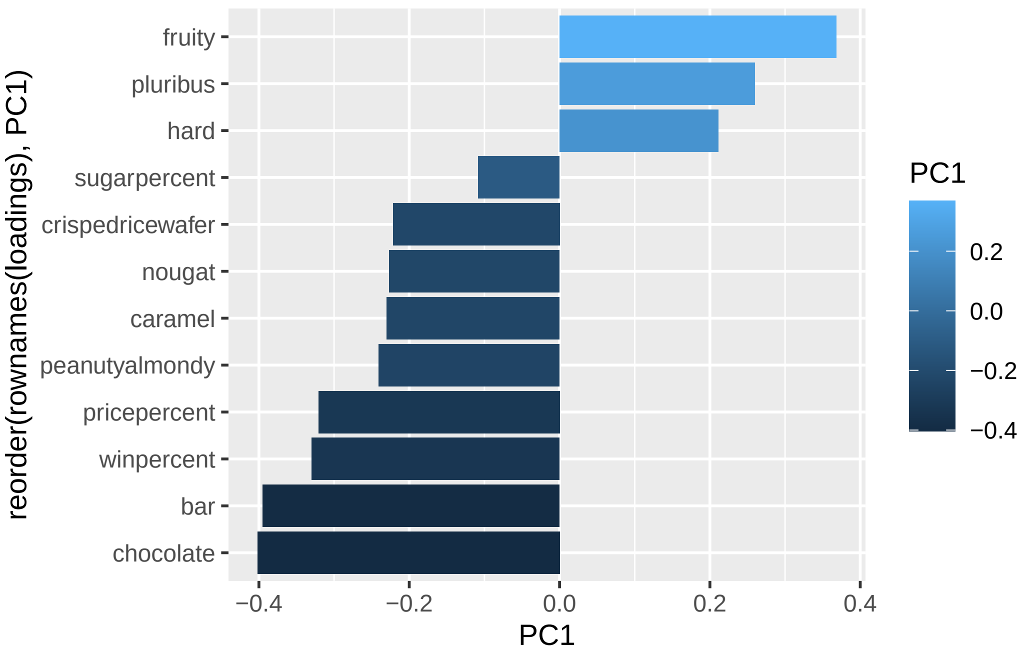
```
Importance of components:
                          PC1    PC2    PC3     PC4    PC5     PC6     PC7
Standard deviation     2.0788 1.1378 1.1092 1.07533 0.9518 0.81923 0.81530
Proportion of Variance 0.3601 0.1079 0.1025 0.09636 0.0755 0.05593 0.05539
Cumulative Proportion  0.3601 0.4680 0.5705 0.66688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

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



How do the original variables (columns) contribute to the new PCs. I will look at the PC1 here.

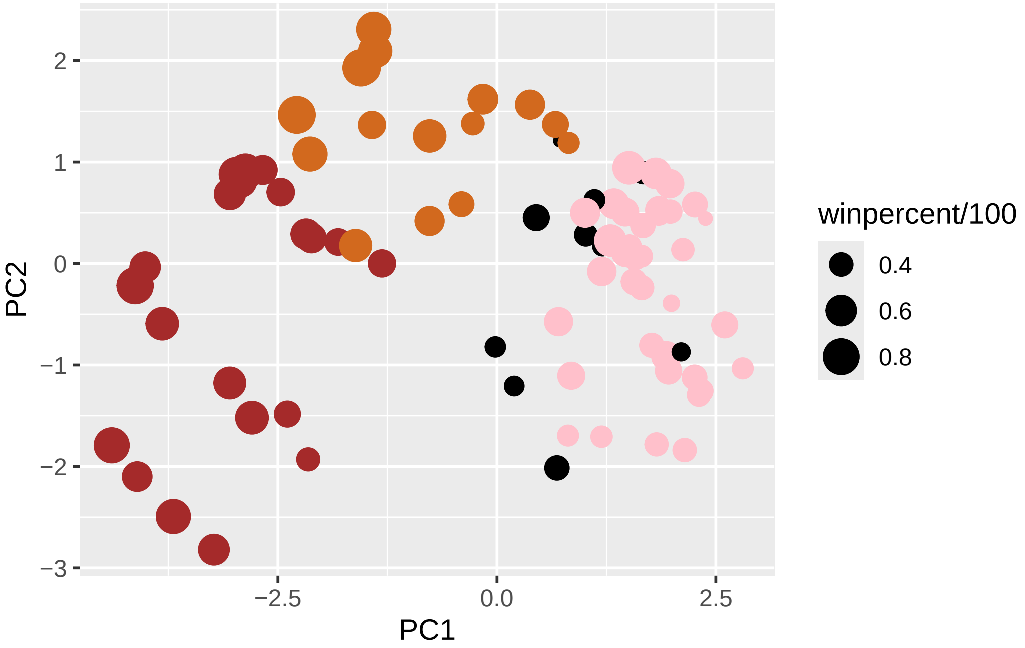
```r
loadings <- as.data.frame(pca$rotation)

ggplot(loadings) +
  aes(PC1, reorder(rownames(loadings), PC1), fill=PC1) +
  geom_col()
```



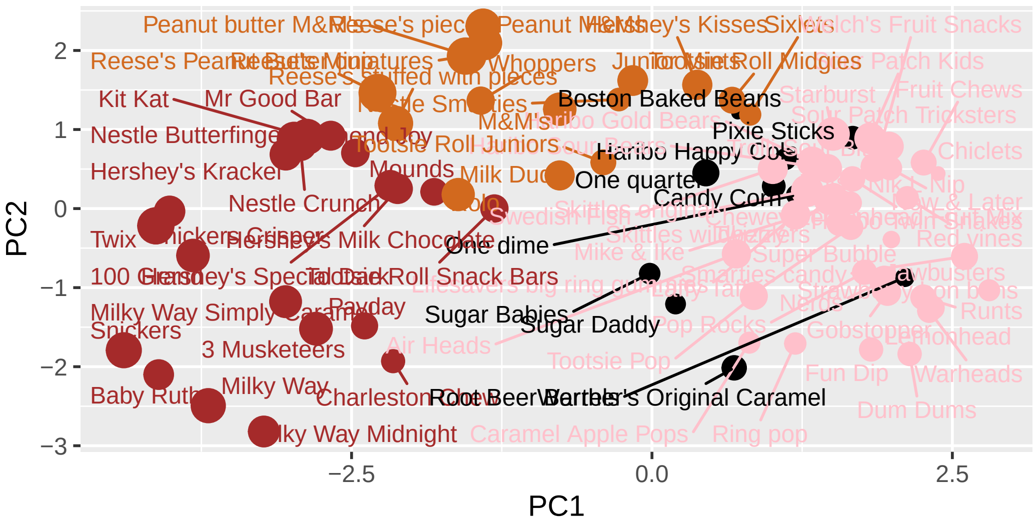
Making nicer plot with ggplot():

```
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 = 49)  +
  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")
```



Halloween Candy PCA Space

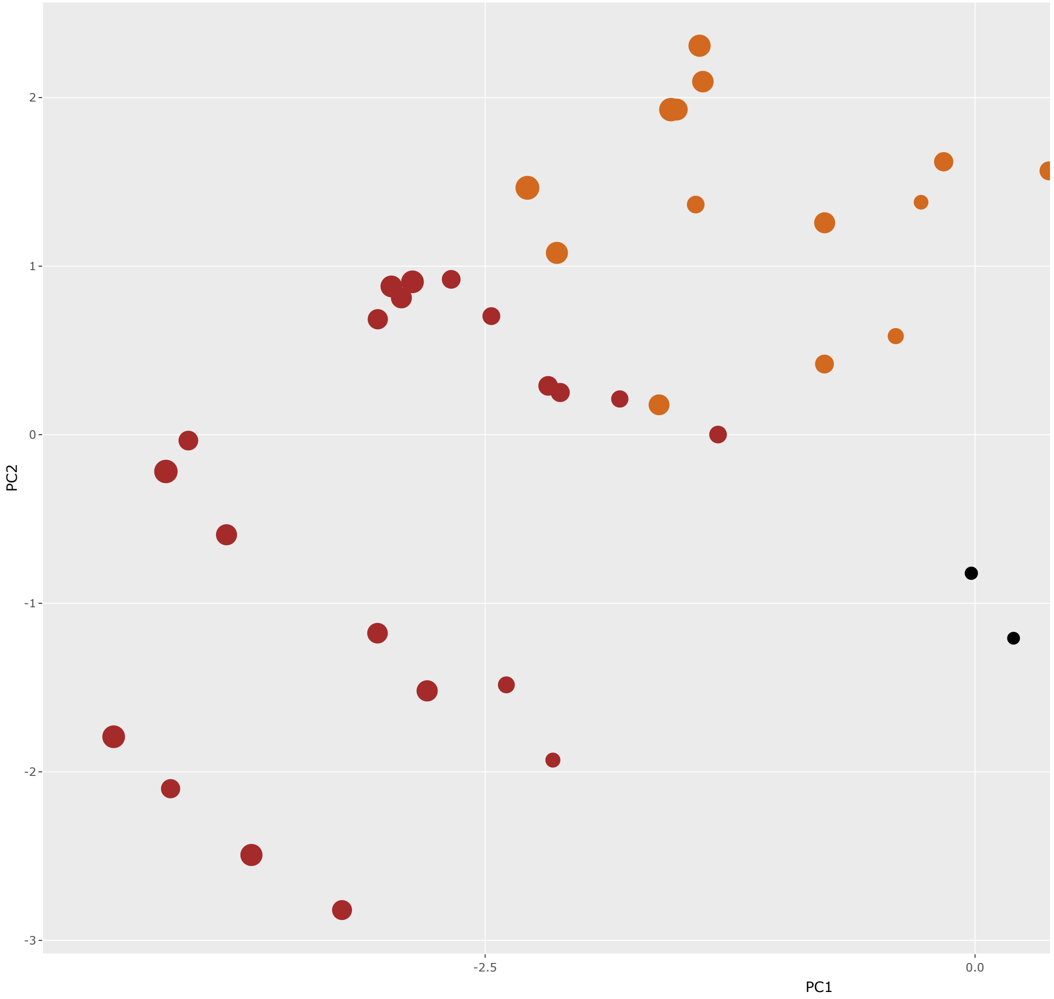
Colored by type: chocolate bar (dark brown), chocolate other (light brown),

Data from 538

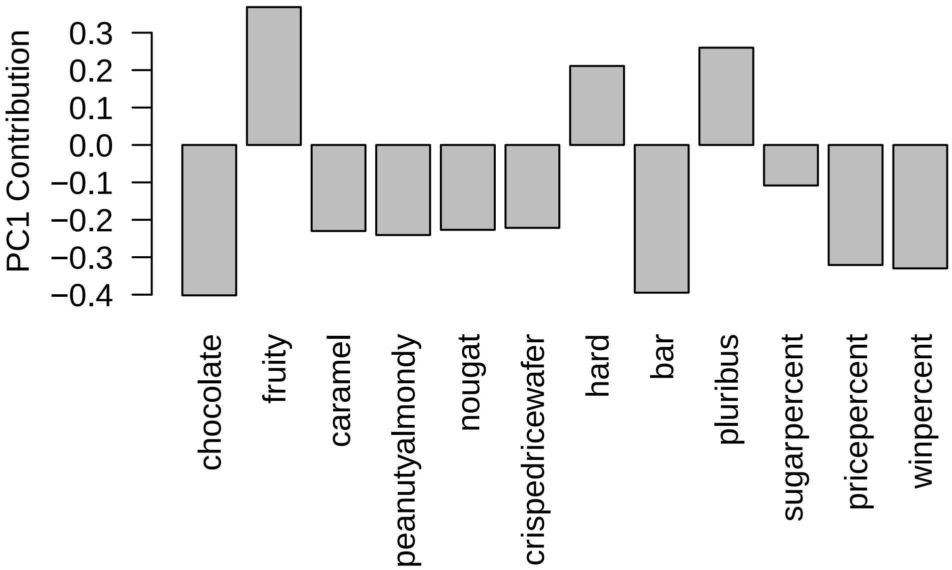
```r
library(plotly)
```

```
Attaching package: 'plotly'

The following object is masked from 'package:ggplot2':

    last_plot

The following object is masked from 'package:stats':

    filter

The following object is masked from 'package:graphics':

    layout
```

```r
ggplotly(p)
```


2
1
0
-1
-2
-3
PC2
-2.5
0.0
PC1

```r
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?

Fruity, pluribus, and hard variables are picked up strongly by PC1 in the positive direction. Yes, this does make sense to me.