Class 10: Halloween Candy Project

Elena

Table of contents

Favourite Candy

Favourite Candy

Importing Candy Data

```
candy_file <- read.csv("candy-data.csv")

candy = data.frame(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?

85 types of candy.

```
nrow(candy)
```

```
[1] 85
```

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

38 fruity candy types.

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

```
[1] 38
```

Q3. What is your favourite candy in the dataset and what is its winpercent value?

Haribo Happy Cola, 34.15896.

```
candy["Haribo Happy Cola", ]$winpercent
```

```
[1] 34.15896
```

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

76.7686.

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

```
[1] 76.7686
```

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

49.6535.

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

```
[1] 49.6535
```

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.

```r
#install.packages("skimr")
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	

```r
skimcandy <- skim(candy)
skimcandy$numeric.mean
```

```
 [1]  0.43529412  0.44705882  0.16470588  0.16470588  0.08235294  0.08235294
 [7]  0.17647059  0.24705882  0.51764706  0.47864705  0.46888235 50.31676381
```

```
#12th value has different scale
#Therefore look at what the 12th variable
skimcandy[12,]
```

Table 3: Data summary

Name	candy
Number of rows	85
Number of columns	12
Column type frequency:	
numeric	1
Group variables	None

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
winpercent	0	1	50.32	14.71	22.45	39.14	47.83	59.86	84.18	

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

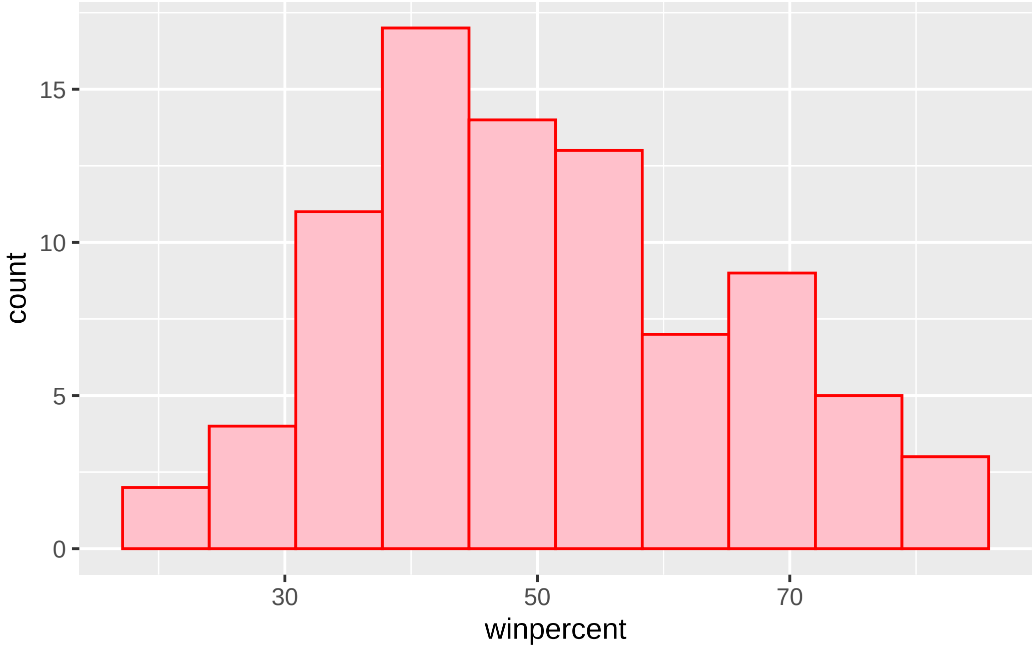
0 means that the candy does not have chocolate in it, whereas 1 indicates that the candy does have chocolate.

```
candy$chocolate
```

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

Q8. Plot a histogram of winpercent values

```
library(ggplot2)
ggplot(candy) + aes(winpercent) + geom_histogram(bins=10, col="red", fill="pink")
```



Q9. Is the distribution of winpercent values symmetrical?

No.

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

Below

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

Chocolate.

```
#Chocolate
chocolate.inds <- as.logical(candy$chocolate)
```

```r
chocolate.win <- candy[chocolate.inds,]$winpercent
mean(chocolate.win)
```

```
[1] 60.92153
```

```r
#Fruit Candy
fruit.inds <- as.logical(candy$fruity)
fruit.win <- candy[fruit.inds,]$winpercent
mean(fruit.win)
```

```
[1] 44.11974
```

Q12. Is this difference statistically significant?

Yes.

```r
t.test(chocolate.win,fruit.win)
```

```

	Welch Two Sample t-test

data:  chocolate.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 
```

Candy Rankings

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

Nik L Nip, Boston Baked Beans, Chiclets, Super Bubble, Jawbusters.

The base R `sort()` and `order()` functions are very useful! Note: `order()` tells how the input would be rearranged.

```r
#Example
x <- c(5,1,2,6)
sort(x)
```

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

```r
#Example
y <- c("barry","alice","chandra")
y
```

```
[1] "barry"   "alice"   "chandra"
```

```r
sort(y)
```

```
[1] "alice"   "barry"   "chandra"
```

```r
order(y)
```

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

```r
inds <- order(candy$winpercent)
inds
```

```
 [1] 45  8 13 73 27 58 72  3 71 20 10 70 60 56 12 51 49 63  9 11 82 31 17 46 15
[26] 50 30 84 22 14 59 76 16 83 81 77 64  4 47 35 18 79 40 75 85 78  6 21  5 68
[51] 32 41 74 36 62 42 23 25  7 19 28 26 66 67 38 24 61 39 57 44 34  1 69  2 48
[76] 43 33 55 37 54 65 29 80 52 53
```

```r
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
WertherÕs 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
WelchÕs 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
HersheyÕs Kisses                    1      0       0              0      0
HersheyÕs 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
HersheyÕs 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
HersheyÕs 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&MÕs                               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&MÕs                 1      0       0              1      0
ReeseÕs stuffed with pieces         1      0       0              1      0
Milky Way                           1      0       1              0      1
ReeseÕs 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
ReeseÕs Miniatures                  1      0       0              1      0
ReeseÕs 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
WertherÕs 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
WelchÕs 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
HersheyÕs Kisses                           0    0   0        1        0.127
HersheyÕs 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
HersheyÕs 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
HersheyÕs 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&MÕs                                      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&MÕs                        0    0   0        1        0.825
ReeseÕs stuffed with pieces                0    0   0        0        0.988
Milky Way                                  0    0   1        0        0.604
ReeseÕs 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
ReeseÕs Miniatures                         0    0   0        0        0.034
ReeseÕs 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
WertherÕs 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
WelchÕs 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
HersheyÕs Kisses                   0.093   55.37545
HersheyÕs 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
HersheyÕs 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
HersheyÕs 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&MÕs                              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&MÕs                0.651   71.46505
ReeseÕs stuffed with pieces        0.651   72.88790
Milky Way                          0.651   73.09956
ReeseÕs pieces                     0.651   73.43499
Snickers                           0.651   76.67378
Kit Kat                            0.511   76.76860
Twix                               0.906   81.64291
ReeseÕs Miniatures                 0.279   81.86626
ReeseÕs Peanut Butter cup          0.651   84.18029
```

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

```
                   chocolate fruity caramel peanutyalmondy nougat
```

```
Nik L Nip                   0      1       0               0      0
Boston Baked Beans          0      0       0               1      0
Chiclets                    0      1       0               0      0
Super Bubble                0      1       0               0      0
Jawbusters                  0      1       0               0      0
                   crispedricewafer hard bar pluribus sugarpercent pricepercent
Nik L Nip                         0    0   0        1        0.197        0.976
Boston Baked Beans                0    0   0        1        0.313        0.511
Chiclets                          0    0   0        1        0.046        0.325
Super Bubble                      0    0   0        0        0.162        0.116
Jawbusters                        0    1   0        1        0.093        0.511
                   winpercent
Nik L Nip            22.44534
Boston Baked Beans   23.41782
Chiclets             24.52499
Super Bubble         27.30386
Jawbusters           28.12744
```

```
#Or, using dplyr...
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 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?

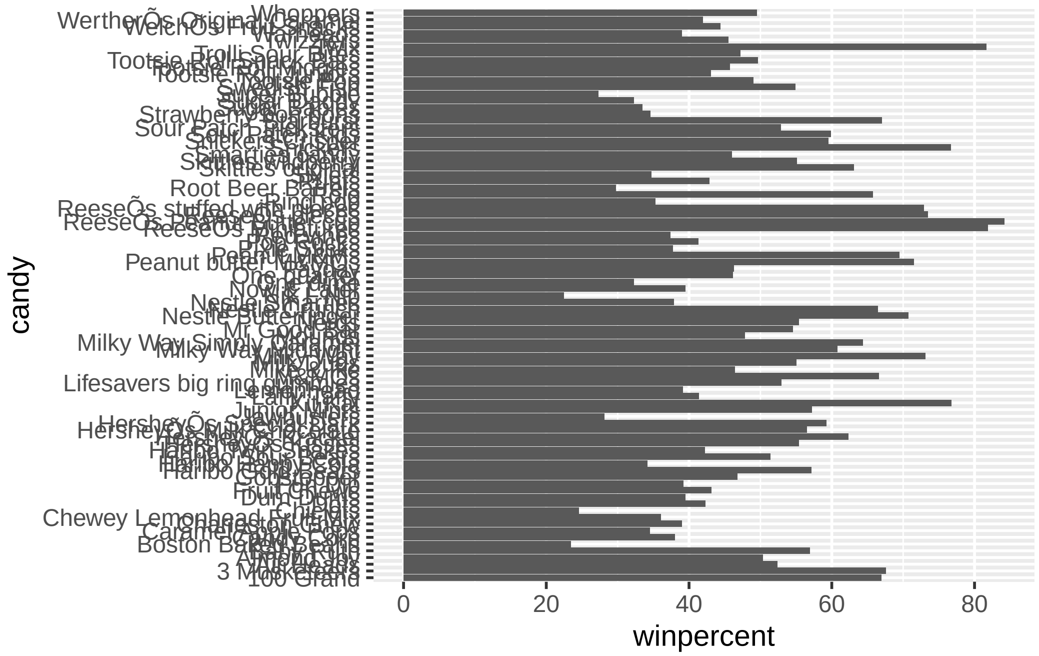
Reese’s Peanut Butter Cup, Reese’s Miniatures, Twix, Kit Kat, Snickers.

```
tail(candy[order(candy$winpercent),], n=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
ReeseÕs Miniatures                1      0       0              1      0
ReeseÕs 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
ReeseÕs Miniatures                       0    0   0        0        0.034
ReeseÕs 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
ReeseÕs Miniatures               0.279   81.86626
ReeseÕs Peanut Butter cup        0.651   84.18029
```

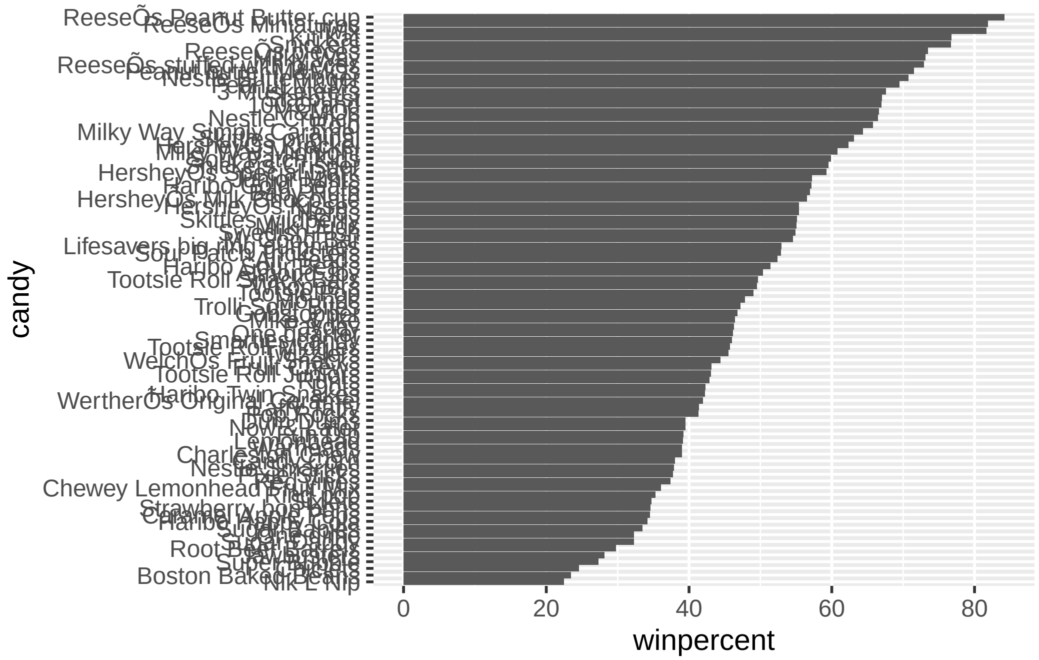
Q15. Make a first barplot of candy ranking based on winpercent values.

```
library(ggplot2)

ggplot(candy) +
  aes(winpercent, rownames(candy)) +
  geom_col() +
  labs(y="candy")
```



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

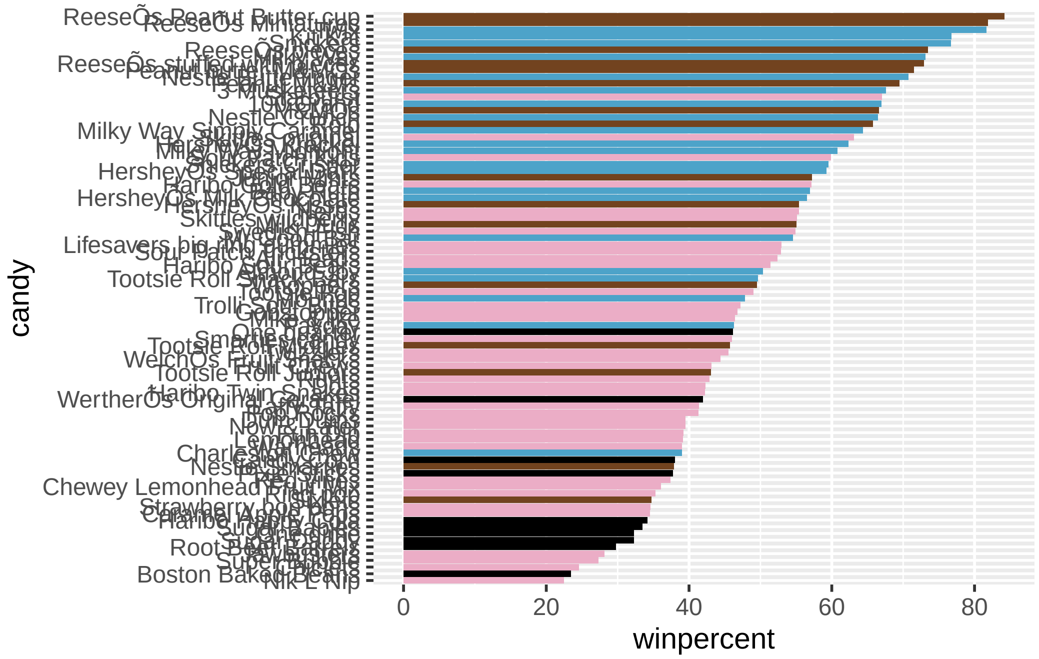
```
ggplot(candy) +
  aes(winpercent, reorder(rownames(candy),winpercent)) +
  geom_col() +
  labs(y="candy")
```



```
#Creating colour vector
#First, create a vector that is all black
my_cols=rep("black", nrow(candy))
#Then, replace choco, bar, fruity with respective colours
my_cols[as.logical(candy$chocolate)] = "#72431F" #brown
my_cols[as.logical(candy$bar)] = "#4DA3C9" #blue
my_cols[as.logical(candy$fruity)] = "#EBADC6" #pink
my_cols
```

```
 [1] "#4DA3C9" "#4DA3C9" "black"   "black"   "#EBADC6" "#4DA3C9" "#4DA3C9"
 [8] "black"   "black"   "#EBADC6" "#4DA3C9" "#EBADC6" "#EBADC6" "#EBADC6"
[15] "#EBADC6" "#EBADC6" "#EBADC6" "#EBADC6" "#EBADC6" "black"   "#EBADC6"
[22] "#EBADC6" "#72431F" "#4DA3C9" "#4DA3C9" "#4DA3C9" "#EBADC6" "#72431F"
[29] "#4DA3C9" "#EBADC6" "#EBADC6" "#EBADC6" "#72431F" "#72431F" "#EBADC6"
[36] "#72431F" "#4DA3C9" "#4DA3C9" "#4DA3C9" "#4DA3C9" "#4DA3C9" "#EBADC6"
[43] "#4DA3C9" "#4DA3C9" "#EBADC6" "#EBADC6" "#4DA3C9" "#72431F" "black"
[50] "#EBADC6" "#EBADC6" "#72431F" "#72431F" "#72431F" "#72431F" "#EBADC6"
[57] "#72431F" "black"   "#EBADC6" "#72431F" "#EBADC6" "#EBADC6" "#72431F"
[64] "#EBADC6" "#4DA3C9" "#4DA3C9" "#EBADC6" "#EBADC6" "#EBADC6" "#EBADC6"
[71] "black"   "black"   "#EBADC6" "#EBADC6" "#EBADC6" "#72431F" "#72431F"
[78] "#4DA3C9" "#EBADC6" "#4DA3C9" "#EBADC6" "#EBADC6" "#EBADC6" "black"
[85] "#72431F"
```

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



```r
ggsave("candybarplot.png")
```

```
Saving 5.5 x 3.5 in image
```

Q17. What is the worst ranked chocolate candy?

Sixlets.

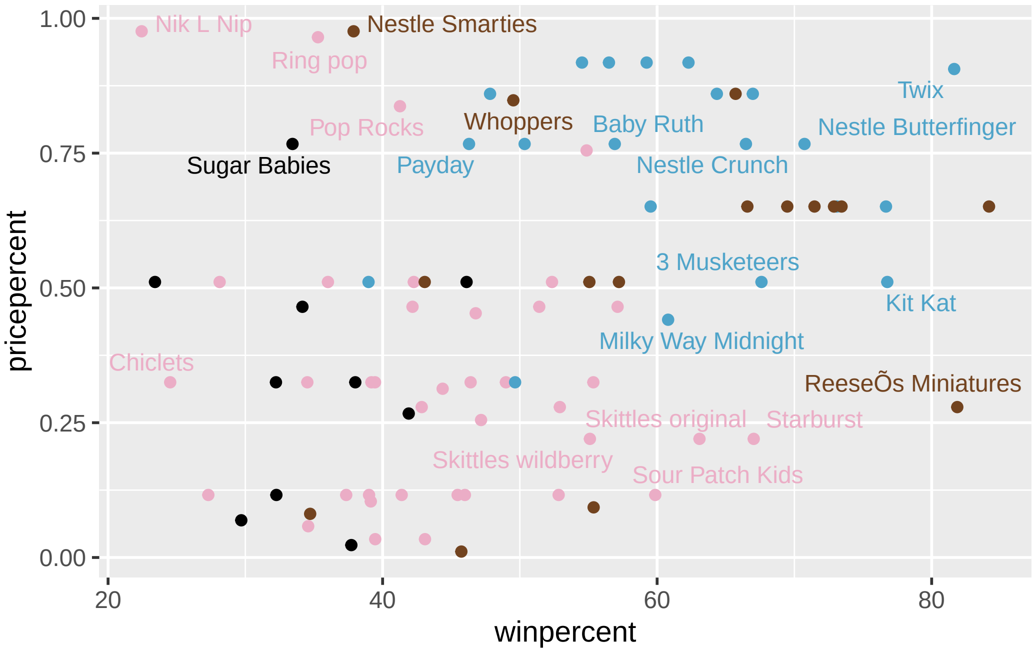
Q18. What is the best ranked fruity candy?

Starburst.

Price

```
#install.packages("ggrepel")
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?

Fruity candy - relatively well liked, and comparatively cheap. Reese's Miniatures is the cheapest of the top five popular candies.

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

Nik L Nip, Nestle Smarties, Ring pop, Hershey's Krackel, Hershey's Milk Chocolate. Nik L Nip is 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
HersheyÕs Krackel               0.918   62.28448
HersheyÕs Milk Chocolate        0.918   56.49050
```

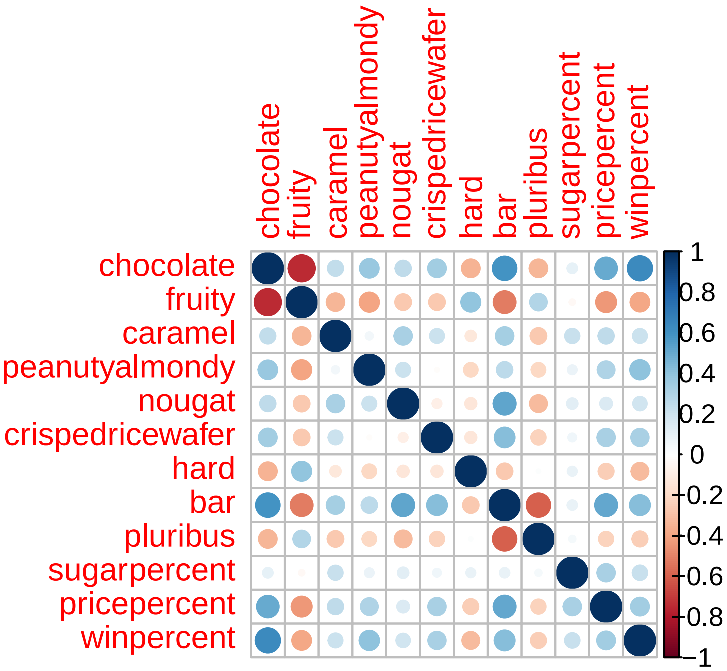
Q21. NA

Correlation

```r
#install.packages("corrplot")
library(corrplot)
```

```
corrplot 0.92 loaded
```

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



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

Chocolate and fruity.

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

Chocolate and bar.

PCA

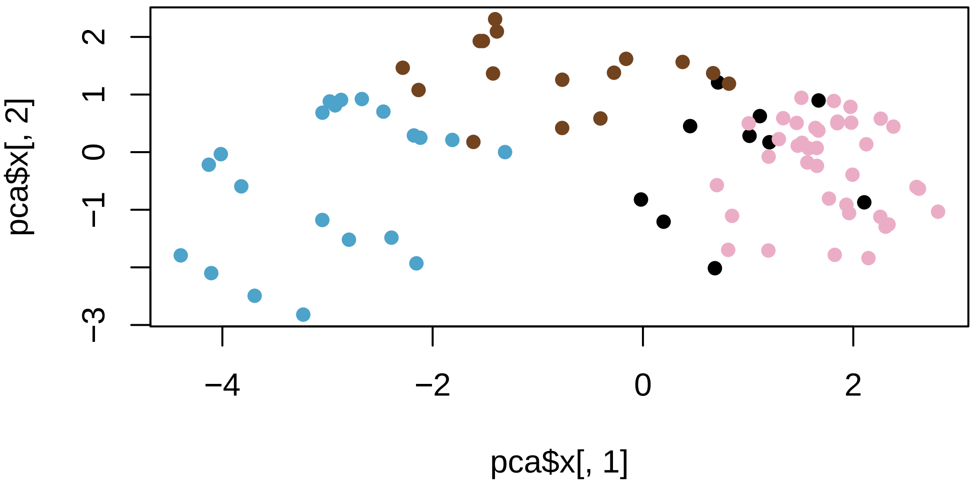
Use `prcomp()` - it has an important argument that is set to `scale=FALSE` by default. In this case, we would want to use `scale=TRUE` due to winpercent being in a different range.

```
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
```

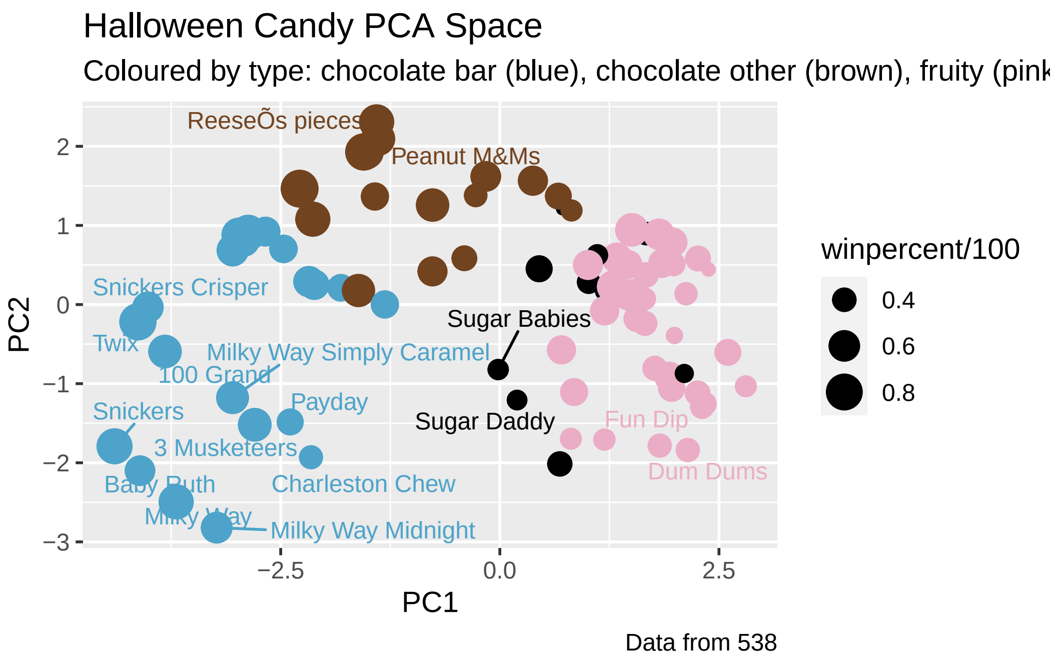
```
plot(pca$x[,1],pca$x[,2],col=my_cols,pch=16)
```



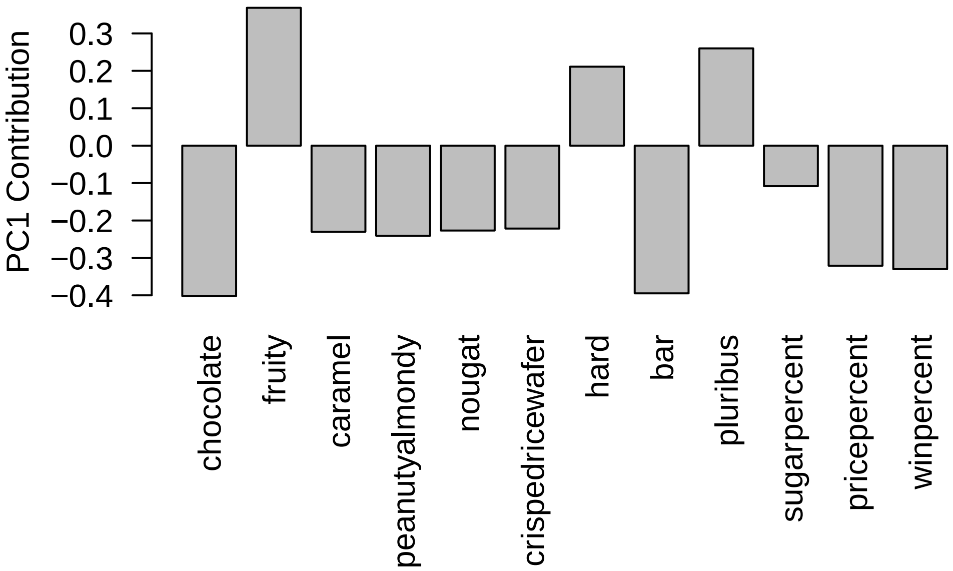
Make ggplot

```
#First, make dataframe in order to make ggplot
#cbind function will add columns with PCA data onto candy dataframe
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) +
        geom_text_repel(col=my_cols, max.overlaps=7, size=3) +
        labs(title="Halloween Candy PCA Space",
       subtitle="Coloured by type: chocolate bar (blue), chocolate other (brown), fruity (
       caption="Data from 538")

p
```

```
Warning: ggrepel: 68 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```



```
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, followed by pluribus and hard. Yes, because we saw that fruity candies are often hard candy that come in a bag or box of multiple candies.