Analytics Consulting Case Study - Süßigkeit Empfehlung

Tihomira Nikolova
tiholova8@gmail.com

Contents

Zielsetzung

Eine Empfehlung für die Eigenschaften einer neuen Süßigkeit

Der Einkauf von dem Supermarkt möchte das Süßwarensortiment (Eigenmarke) erweitern. Hierzu soll eine neue Süßigkeit kreiert werden. Jedoch besteht innerhalb des betreffenden Projektteams noch Uneinigkeit über die Charak-teristika der neuen Süßigkeit. Während die Einen beispielsweise eine Keks-basierte Süßigkeit bevorzugen, favorisieren Andere eine Fruchtgummivariation. Daher hat der Bereichsvorstand beschlossen ein Markt-forschungsunternehmen zu beauftragen, um die Beliebtheit, der am Markt erhältlichen Süßwaren, zu er-mitteln. Die Ergebnisse der Marktforschung liegen nun vor und Sie wurden beauftragt eine Analyse der Daten durchzuführen. Ziel ist es, die Auswirkungen der Charakteristika von Süßwaren auf deren Beliebtheit zu analysieren und auf Basis dieser Analyse eine Empfehlung für die Eigenschaften einer neuen Süßigkeit abzugeben.

Daten Den Datensatz (inkl. einer kurzen Beschreibung) finden Sie unter: [https://github.com/fivethirtyeight/data/tree/master/candy-power-ranking] Hierbei handelt es sich um einen Datensatz von FiveThirtyEight, der unter der Creative Commons Attribution 4.0 International license [https://creativecommons.org/licenses/by/4.0/] steht.

Data

This folder contains the data behind the story The Ultimate Halloween Candy Power Ranking.

candy-data.csv includes attributes for each candy along with its ranking. For binary variables, 1 means yes, 0 means no.

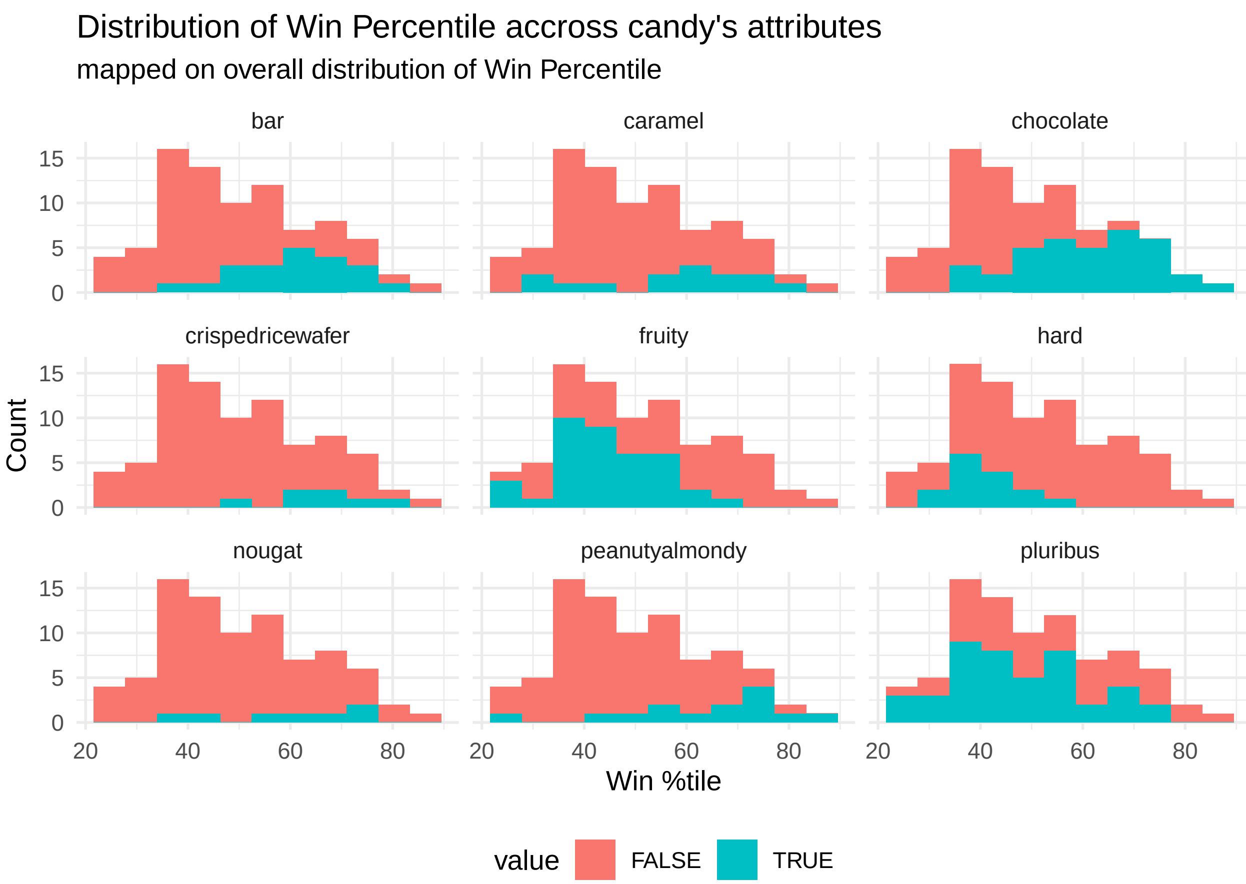
The data contains the following fields:

Header Description
chocolate Does it contain chocolate?
fruity Is it fruit flavored?
caramel Is there caramel in the candy?
peanutalmondy Does it contain peanuts, peanut butter or almonds?
nougat Does it contain nougat?
crispedricewafer Does it contain crisped rice, wafers, or a cookie component?
hard Is it a hard candy?
bar Is it a candy bar?
pluribus Is it one of many candies in a bag or box?
sugarpercent The percentile of sugar it falls under within the data set.
pricepercent The unit price percentile compared to the rest of the set.
winpercent The overall win percentage according to 269,000 matchups.

Wir werden uns die Daten ansehen, mit denen wir arbeiten.

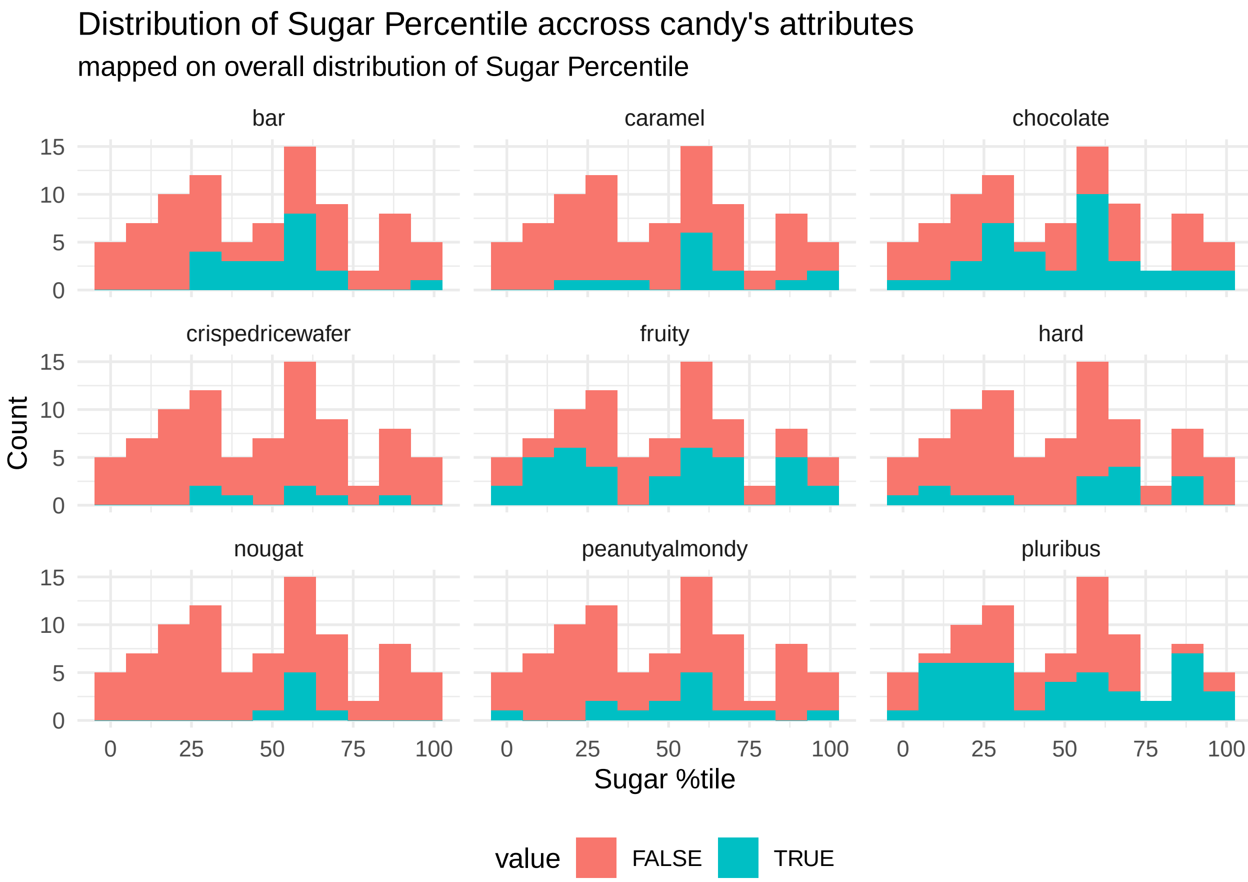
```
## tibble [85 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ competitorname  : chr [1:85] "100 Grand" "3 Musketeers" "One dime" "One quarter" ...
##  $ chocolate       : logi [1:85] TRUE TRUE FALSE FALSE FALSE TRUE ...
##  $ fruity          : logi [1:85] FALSE FALSE FALSE FALSE TRUE FALSE ...
##  $ caramel         : logi [1:85] TRUE FALSE FALSE FALSE FALSE FALSE ...
##  $ peanutyalmondy  : logi [1:85] FALSE FALSE FALSE FALSE FALSE TRUE ...
##  $ nougat          : logi [1:85] FALSE TRUE FALSE FALSE FALSE FALSE ...
##  $ crispedricewafer: logi [1:85] TRUE FALSE FALSE FALSE FALSE FALSE ...
##  $ hard            : logi [1:85] FALSE FALSE FALSE FALSE FALSE FALSE ...
##  $ bar             : logi [1:85] TRUE TRUE FALSE FALSE FALSE TRUE ...
##  $ pluribus        : logi [1:85] FALSE FALSE FALSE FALSE FALSE FALSE ...
##  $ sugarpercent    : num [1:85] 0.732 0.604 0.011 0.011 0.906 ...
##  $ pricepercent    : num [1:85] 0.86 0.511 0.116 0.511 0.511 ...
##  $ winpercent      : num [1:85] 67 67.6 32.3 46.1 52.3 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   competitorname = col_character(),
##   ..   chocolate = col_double(),
##   ..   fruity = col_double(),
##   ..   caramel = col_double(),
##   ..   peanutyalmondy = col_double(),
##   ..   nougat = col_double(),
##   ..   crispedricewafer = col_double(),
##   ..   hard = col_double(),
##   ..   bar = col_double(),
##   ..   pluribus = col_double(),
##   ..   sugarpercent = col_double(),
##   ..   pricepercent = col_double(),
##   ..   winpercent = col_double()
##   .. )
```

EDE Histograms



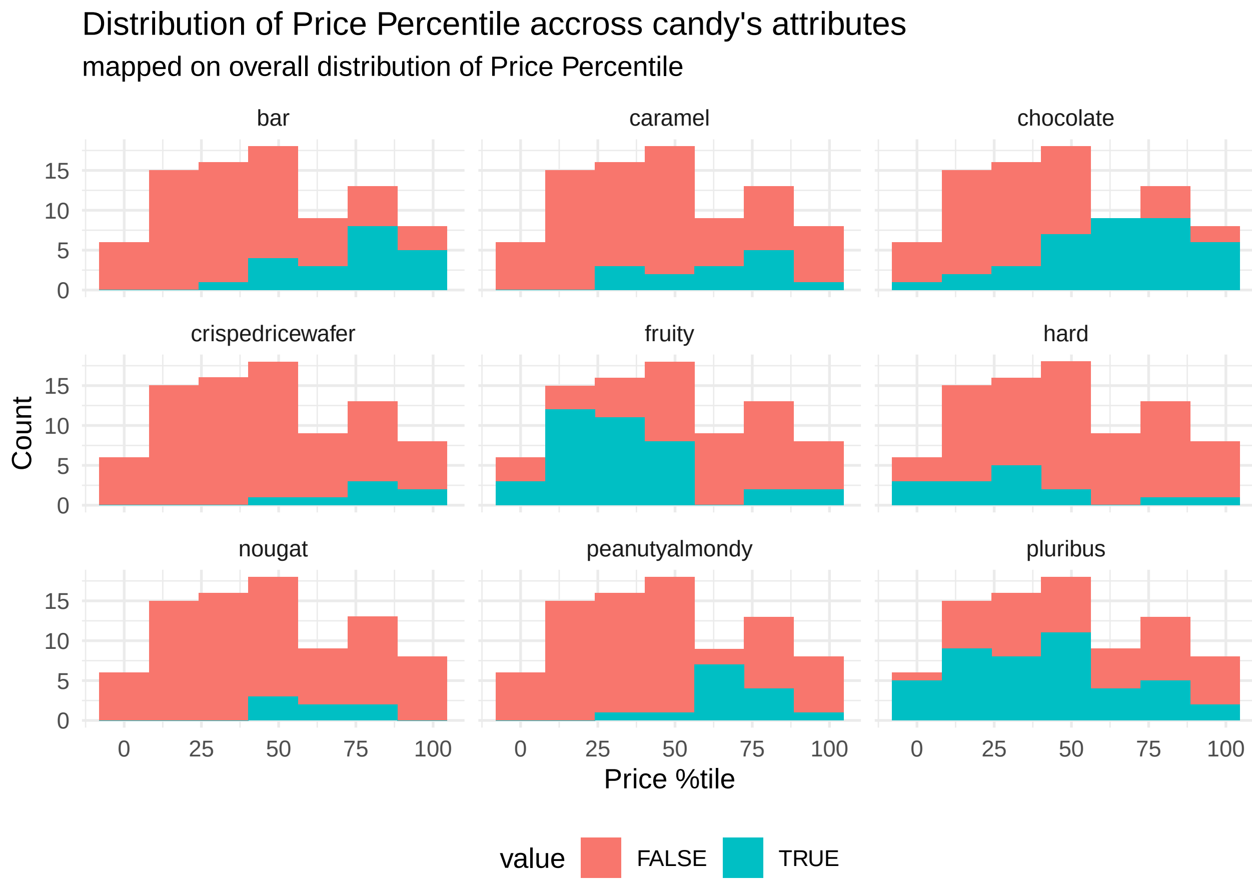
Hier sehen wir die Verteilung jeder Variablen, die über der Verteilung von **winpercent** abgebildet wird Auf den ersten Blick scheint das **winpercent** normal verteilt zu sein, aber ich werde später einen zusätzlichen Test durchführen, um diesen Verdacht zu unterstützen. Die Wahl des Binning kann irreführend sein. Wir sehen auch, dass **chocolate**, **bar**, **crispedricewafer**, **peanutyalmondy** einen hohen **winpercent** Score haben. **caramel** ist interessant, da es sowohl einen hohen als auch einen niedrigen **winpercent** zu haben scheint.
fruity, **pluribus**, **hard** liegt offensichtlich am unteren Ende der Winpercent-Verteilung.

Here we see the distribution of every variable mapped on top of the distribution of **winpercent**. At first glance, it seems that the **winpercent** is normally distributed but I'll perform an additional test later to support this suspicion.The choice of binning may be misleading. We see also that **chocolate**, **bar** , **crispedricewafer**, **peanutyalmondy** have a high **winpercent** score. **caramel** is interesting as it appears to have both high and low **winpercent**.
fruity ,**pluribus** , **hard** are obviously on the lower end of the winpercent distribution.



Süßigkeiten mit diesen Zutaten neigen dazu, in der Süße zu variieren. Mit Ausnahme von **nougat** konzentriert es sich auf das 60. Perzentil.

Candies with these ingredients tend to vary in sweetness. except for **nougat**, it is focused on the 60th percentile.

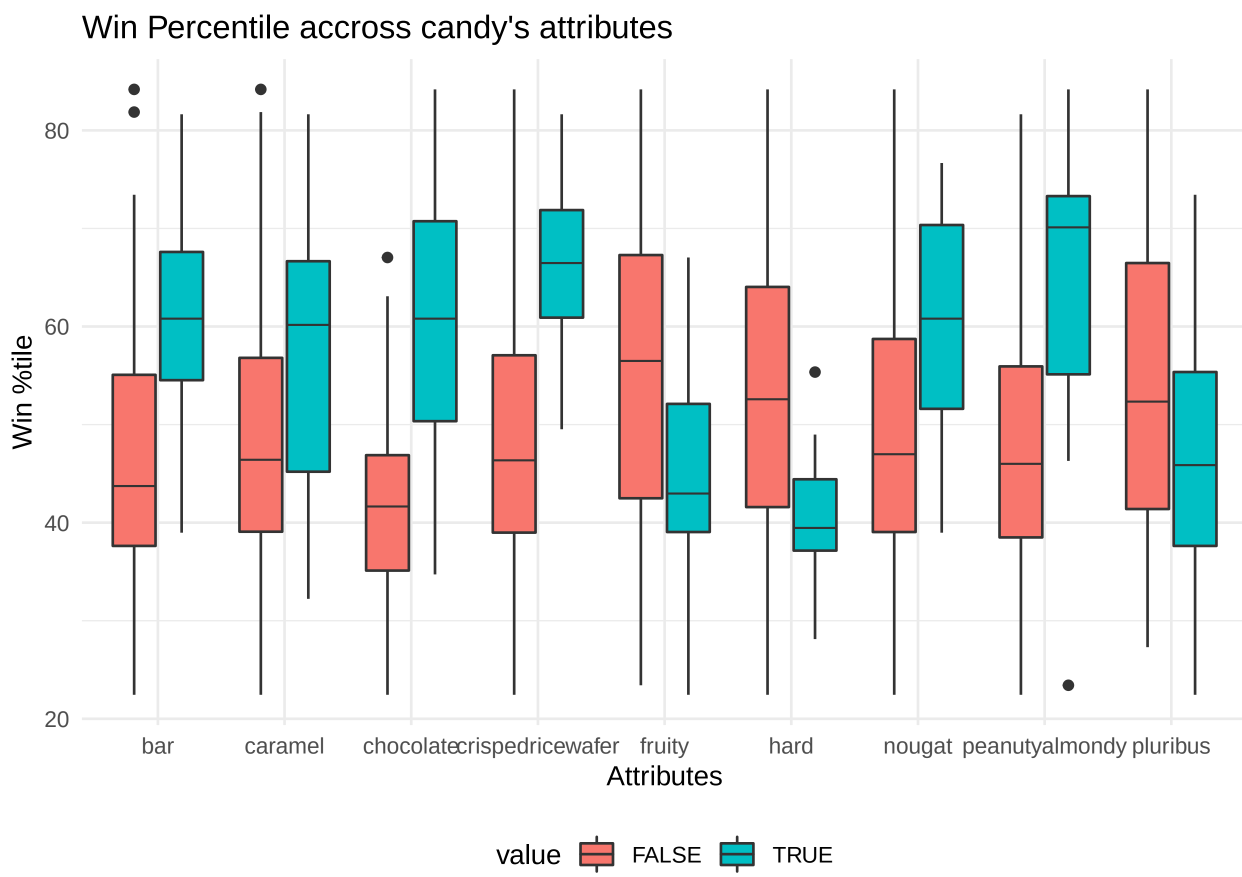


bar, **caramel**, **chocolate**, **crispedricewafer**, **peanutyalmondy** besetzen tendenziell die höheren Preisklassen, wobei **fruity** und **hard** die niedrigeren.

bar, **caramel**, **chocolate**, **crispedricewafer**, **peanutyalmondy** tend to occupy the higher price ranks where **fruity** and **hard** the lower ranks.

EDE Boxplots

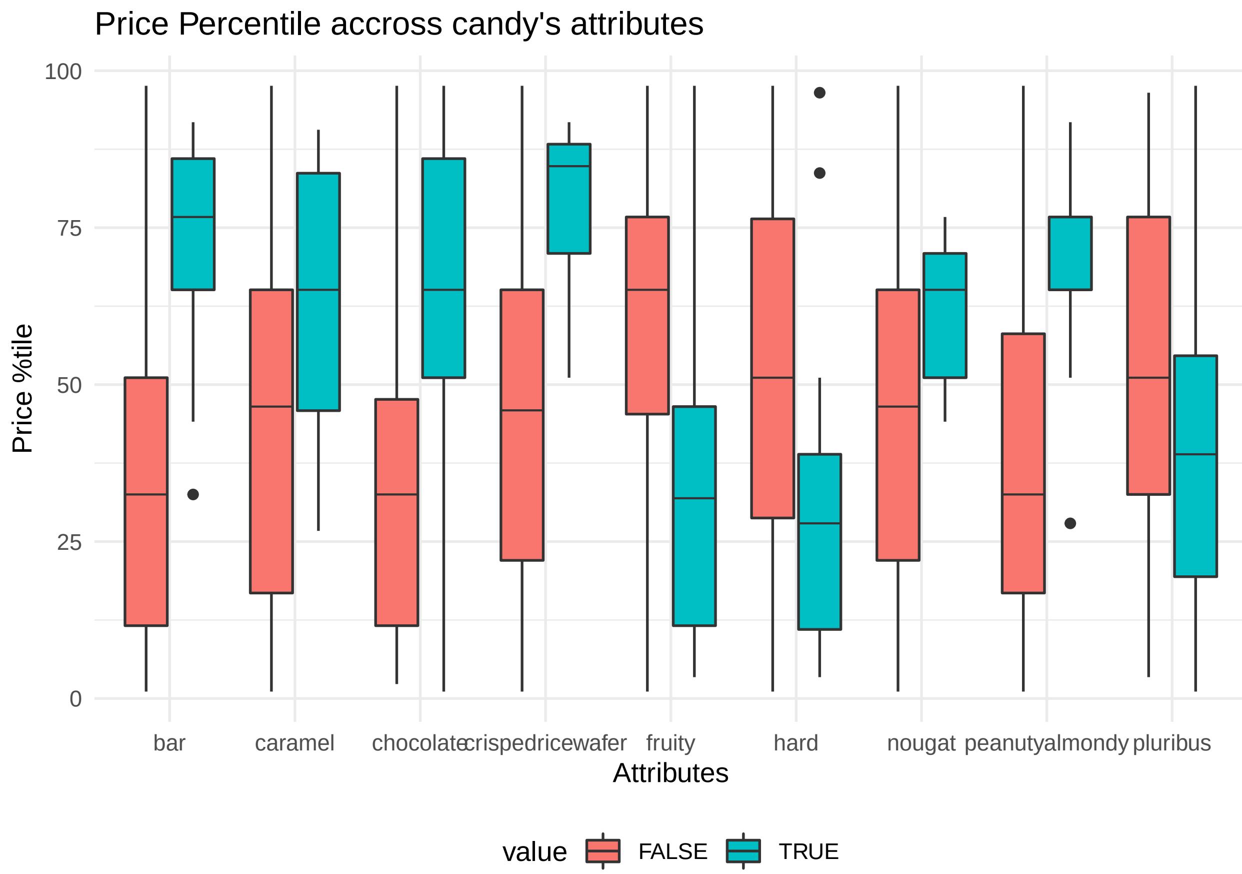
Hier vergleiche ich visuell das Vorhandensein und Fehlen eines Attributs in einer Süßigkeit.
Here I visually compare the presence and absence of an attribute in a candy.



Was wir hier feststellen können, ist, dass das Vorhandensein von fruchtig und hart einen geringeren Winteranteil aufweist als das Fehlen von Süßigkeiten. Die längeren Schwänze deuten auch auf eine hohe Variabilität hin. Ich werde die Outliers extrahieren, damit ich sie mir ansehen kann. Warum unterscheiden sie sich von den anderen?

What we can notice here is that the presence of fruity and hard have lower winpercent compared to the absence of them in candy. The longer tails also suggest high variability. I'll extract the outliers so I can take a look at them and look for an answer Why do they differ from the rest?

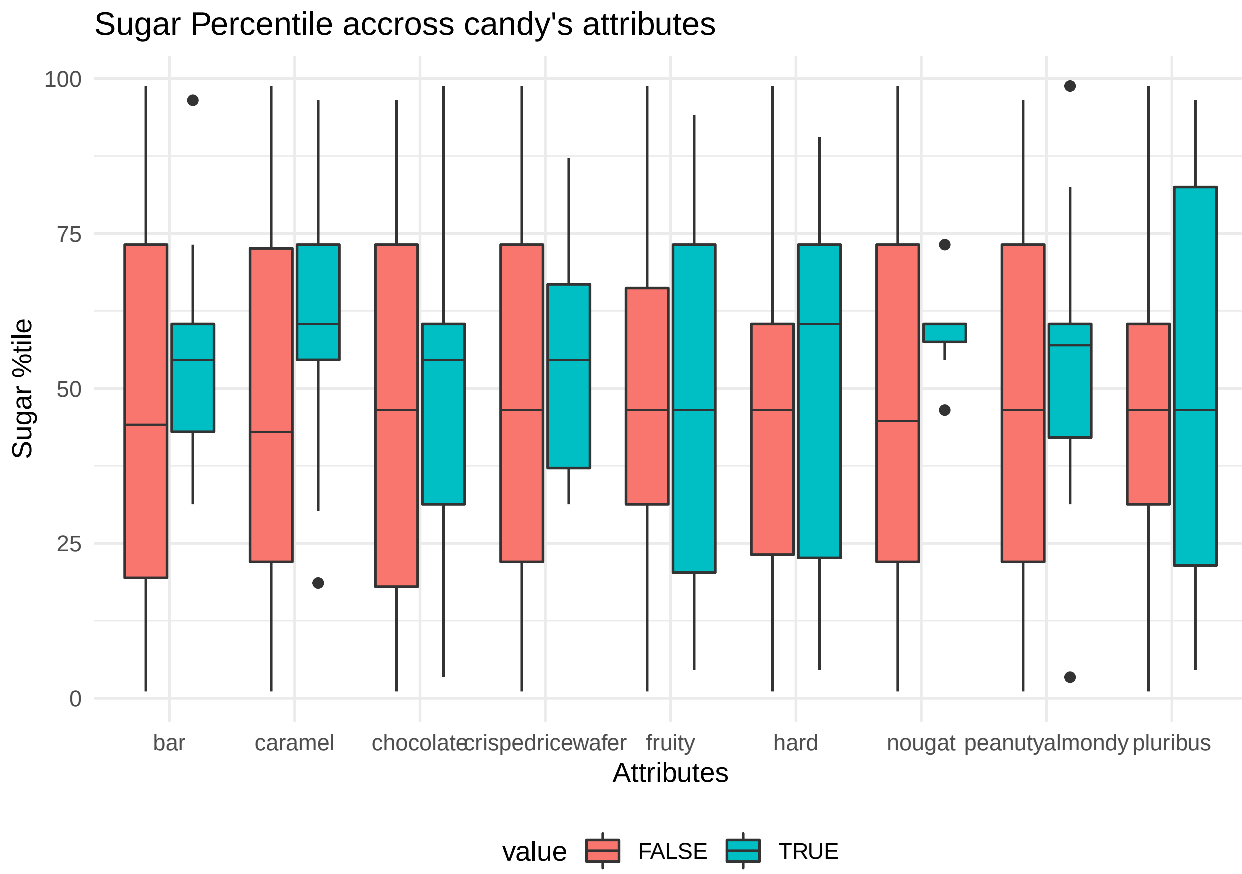
competitorname	chocolate	fruity	caramel	peanutyalmondy	nougat	crispedricewafer	hard	bar	pluribus	sugarpercent	pricepercent	winpercent
Starburst	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	0.151	0.220	67.03763
ReeseÕs Peanut Butter cup	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	0.720	0.651	84.18029
Nerds	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	0.848	0.325	55.35405
ReeseÕs Miniatures	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	0.034	0.279	81.86626
ReeseÕs Peanut Butter cup	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	0.720	0.651	84.18029
Boston Baked Beans	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	0.313	0.511	23.41782



Ich werde die Outliers extrahieren, damit ich sie mir ansehen kann. Warum unterscheiden sie sich von den anderen?
Was wir hier bemerken ist, dass der picepercent springt, wenn es bar oder caramel oder chocolate oder crispedricewafer oder peanutyalmondy gibt.

What we notice here is that the picepercent jumps when there is bar or caramel or chocolate or crispedricewafer or peanutyalmondy. I'll extract the outliers so I can take a look at them and look for an answer Why do they differ from the rest?

competitorname	chocolate	fruity	caramel	peanutyalmondy	nougat	crispedricewafer	hard	bar	pluribus	sugarpercent	pricepercent	winpercent
ReeseÕs Miniatures	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	0.034	0.279	81.86626
Pop Rocks	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	0.604	0.837	41.26551
Ring pop	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	0.732	0.965	35.29076
Tootsie Roll Snack Bars	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	0.465	0.325	49.65350



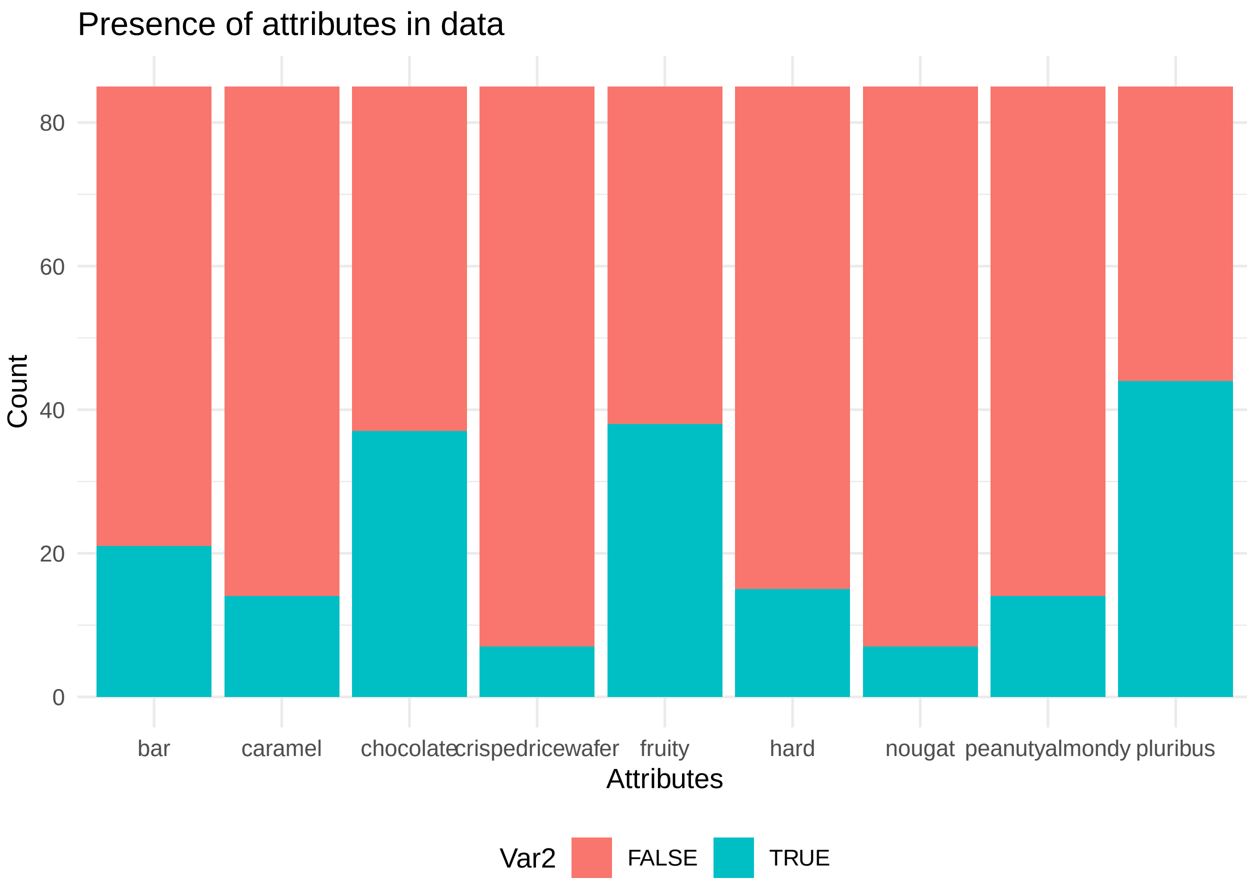
Insgesamt ist der durchschnittliche Zuckeranteil für ein vorhandenes oder fehlendes Attribut in Süßigkeiten gleich. Aber natürlich kann ein `t.test()` helfen, meinen Verdacht zu stützen. Was auffällt, ist **nougat**, **peanutyalmondy caramel**. Ihre Anwesenheit in Süßigkeiten bringt die Süßigkeiten auf einen höheren Rang. Ich werde noch einmal einen Blick auf die Outliers werfen.

Overall the average sugarpercent for a present or absent attribute in candy is the same. But of course a `t.test()` may help to support my suspicion. What stands out is **nougat**, **peanutyalmondy**, and **caramel**. Their presence in candy puts the candy on a higher rank. again, I'll take a look at the outliers.

competitorname	chocolate	fruity	caramel	peanutyalmondy	nougat	crispedricewafer	hard	bar	pluribus	sugarpercent	pricepercent	winpercent
WertherÕs Original Caramel	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	0.186	0.267	41.90431
ReeseÕs Miniatures	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	0.034	0.279	81.86626
ReeseÕs stuffed with pieces	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	0.988	0.651	72.88790
Milky Way Midnight	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	0.732	0.441	60.80070
Payday	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	0.465	0.767	46.29660
Milky Way Simply Caramel	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	0.965	0.860	64.35334

Barplot

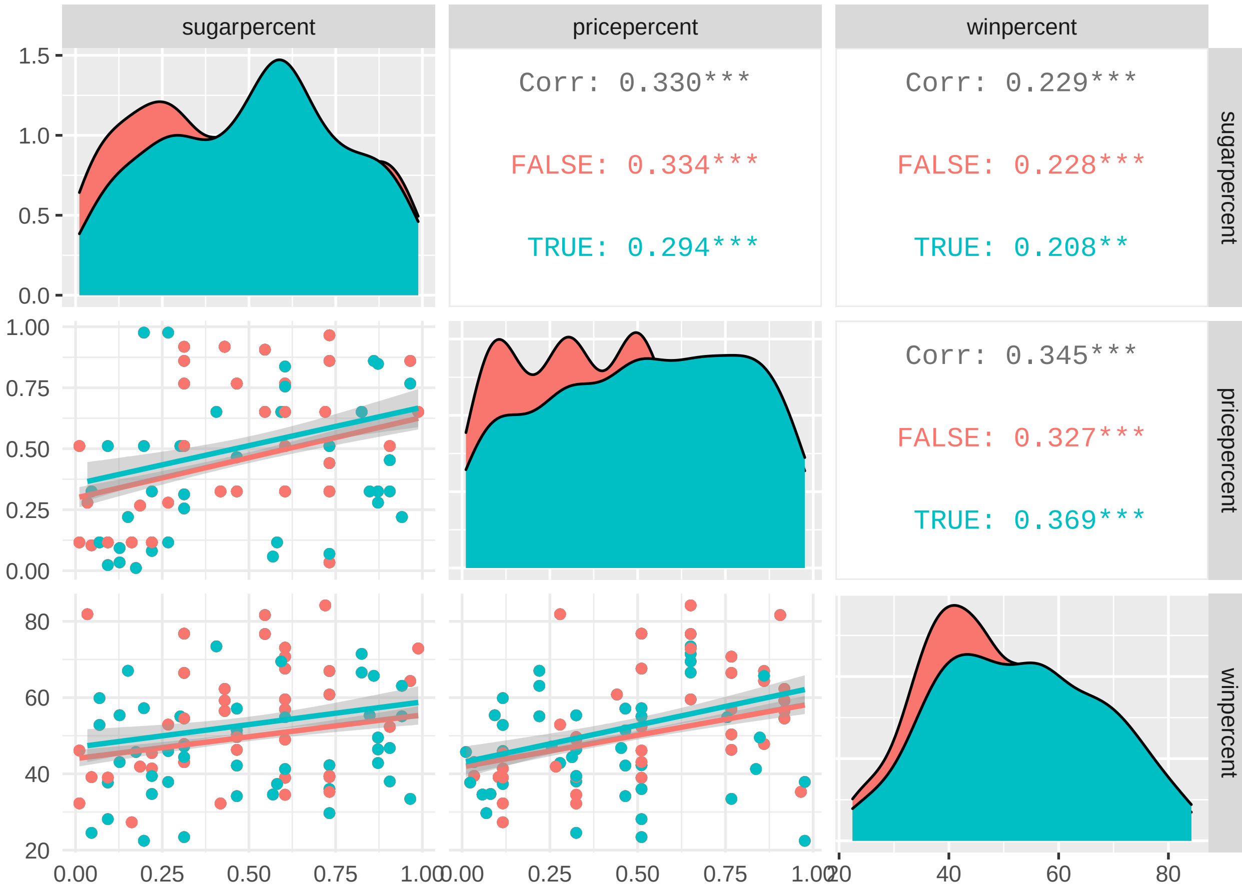
Overall data barplot



Hier sehen wir, dass das Vorhandensein bis Fehlen eines Attributs in einer Süßigkeit im Datensatz. Insgesamt gibt es nicht viele Süßigkeiten mit **crispedricewafer** und **nougat**.

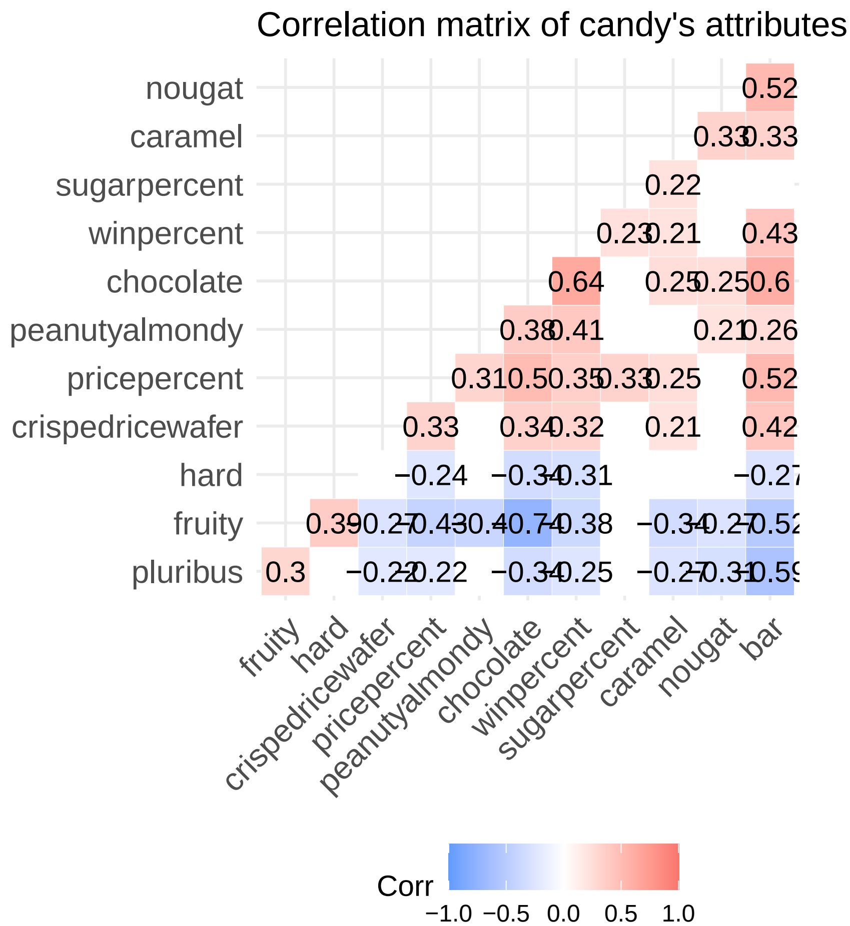
Here we see that the presence to absence of an attribute in a candy in the dataset. Overall there are not many sweets with **crispedricewafer** and **nougat**.

Correlation Exploration



Hier sehen wir, dass die Korrelationen b / n der numerischen Variablen als signifikant markiert sind: zwischen ± 0,30 und ± 0,49 = medium correlation unter ± 0,29 = low correlation
Wenn ich die Variablen in das Modell einführe, erhalte ich möglicherweise irreführende Ergebnisse. Ich werde die Variablen höchstwahrscheinlich auf einen bestimmten Schwellenwert setzen - 50% oder ich würde den Mittelwert oder Median verwenden. Die Dichteverteilung zeigt uns hier die Normalität der Daten. Aber wie ich bereits sagte, werde ich einen statistischen Test verwenden - einen QQ-Plot oder einen `Shapiro-Wilk()` Test, um meine Aussagen zu stützen.

Here we see that the correlations b/n the numerical variables are marked as significant:
between ± 0.30 and ± 0.49 = medium correlation
below ± 0.29 = low correlation
If I introduce the variables into the model I may obtain misleading results. I will most likely bin the variables on some threshold - 50% or I would use the mean or median. Density distribution here shows us the normality of the data. But as I said earlier I will use a statistical test - a QQ plot or `Shapiro-Wilk()` test to support my statements.



In dieser Korrelationsmatrix werden nur signifikante Korrelationen angezeigt. zwischen ± 0,50 und ± 1 = starke Korrelation zwischen ± 0,30 und ± 0,49 = mittlere Korrelation unter ± 0,29 = geringe Korrelation

Was wir feststellen können, ist die negative Korrelation zwischen Schokolade und Frucht. Dies bedeutet, dass **chocolate** häufiger nicht mit **fruity** kombiniert wird. Wir sehen auch eine starke Korrelation zwischen b / n **chocolate** und **winpercent**. Das mag unser stärkstes Attribut für einen hohen **winpercent** sein. . . . Wenn ich die Variablen in das Modell einführe, kann ich Multikollinearität einführen und irreführende Ergebnisse erhalten.

In this correlation matrix, only significant correlations are displayed. between ± 0.50 and ± 1 = strong correlation between ± 0.30 and ± 0.49 = medium correlation
below ± 0.29 = low correlation

What we can notice is the negative correlation b/n **chocolate** and **fruity**. This means that **chocolate** is more often not combined with **fruity**.
We also see a strong correlation b/n **chocolate** and **winpercent**. That may be our strongest attribute for a high **winpercent**. . . . . . .
If I introduce the variables into the model I may introduce multicollinearity and obtain misleading results.
Let's see the candies with **fruity** and **chocolate**.

competitorname	chocolate	fruity	caramel	peanutyalmondy	nougat	crispedricewafer	hard	bar	pluribus	sugarpercent	pricepercent	winpercent
Tootsie Pop	1	1	0	0	0	0	1	0	0	0.604	0.325	48.98265

There is only one candy having both **chocolate** and **fruity** and it has win%tile 49.

First Multiple regression

All attributes as predictors

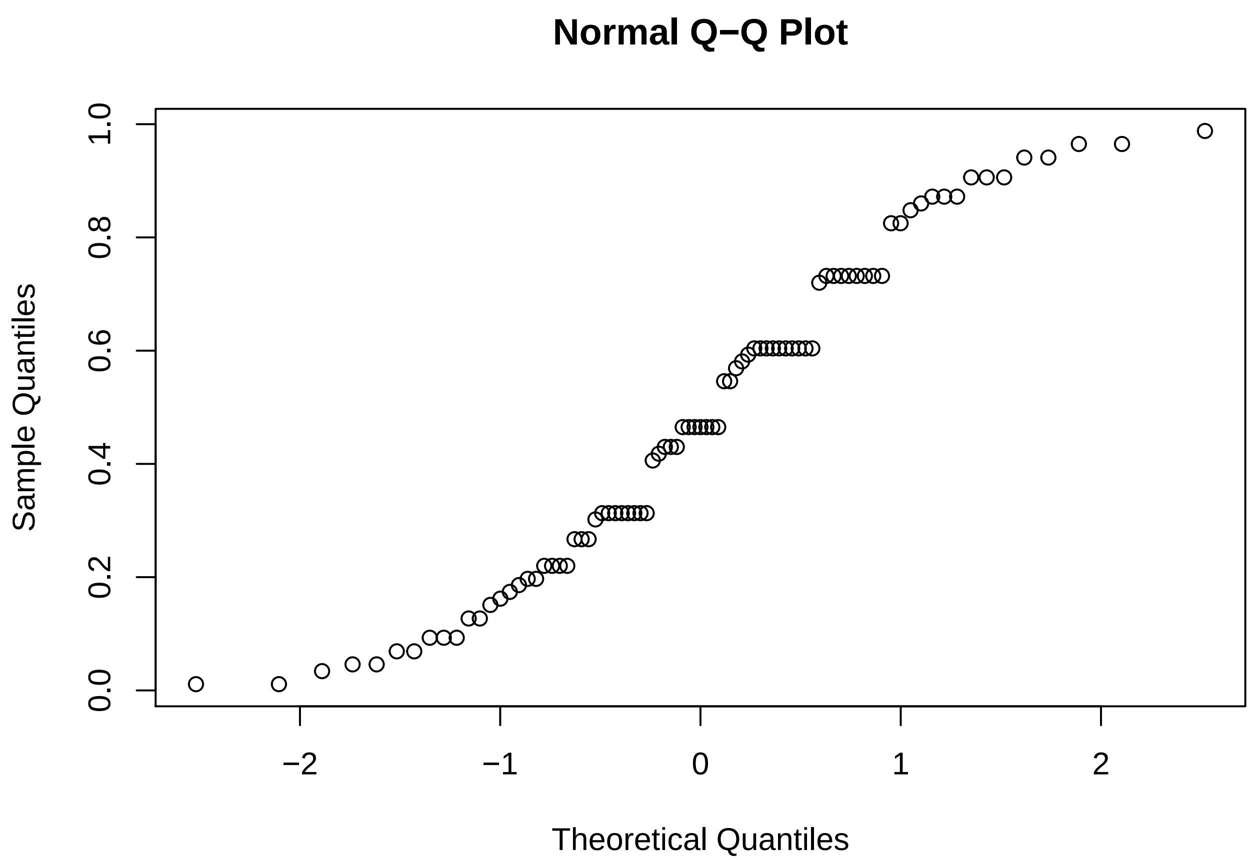
```
##
## Call:
## lm(formula = winpercent ~ ., data = candy[, -1])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.2244  -6.6247   0.1986   6.8420  23.8680
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           34.5340     4.3199   7.994 1.44e-11 ***
## chocolateTRUE         19.7481     3.8987   5.065 2.96e-06 ***
## fruityTRUE             9.4223     3.7630   2.504  0.01452 *
## caramelTRUE            2.2245     3.6574   0.608  0.54493
## peanutyalmondyTRUE    10.0707     3.6158   2.785  0.00681 **
## nougatTRUE             0.8043     5.7164   0.141  0.88849
## crispedricewaferTRUE   8.9190     5.2679   1.693  0.09470 .
## hardTRUE              -6.1653     3.4551  -1.784  0.07852 .
## barTRUE                0.4415     5.0611   0.087  0.93072
## pluribusTRUE          -0.8545     3.0401  -0.281  0.77945
## sugarpercent           9.0868     4.6595   1.950  0.05500 .
## pricepercent          -5.9284     5.5132  -1.075  0.28578
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.7 on 73 degrees of freedom
## Multiple R-squared:  0.5402, Adjusted R-squared:  0.4709
## F-statistic: 7.797 on 11 and 73 DF,  p-value: 9.504e-09
```

From the output of the row model, we may falsely suggest that **chocolate** and **fruity** increase **winpercent**. There is a correlation b/n those two variables and only one observation having both **chocolate** and **fruity**. I will remove this one particular observation from the model.

```
##
## Call:
## lm(formula = winpercent ~ ., data = candy[-indx, -1])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.8677  -5.7754   0.1916   6.6352  24.0721
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           33.7877     4.3390   7.787 3.83e-11 ***
## chocolateTRUE         21.7913     4.1913   5.199 1.80e-06 ***
## fruityTRUE            10.6697     3.8685   2.758  0.00736 **
## caramelTRUE            2.1560     3.6412   0.592  0.55562
## peanutyalmondyTRUE     9.8202     3.6047   2.724  0.00808 **
## nougatTRUE             1.0194     5.6929   0.179  0.85839
```

```
## crispedricewaferTRUE   8.7640     5.2454   1.671  0.09910 .
## hardTRUE              -5.2592     3.5103  -1.498  0.13845
## barTRUE               -0.5978     5.1020  -0.117  0.90705
## pluribusTRUE          -1.5342     3.0718  -0.499  0.61899
## sugarpercent           9.3021     4.6413   2.004  0.04881 *
## pricepercent          -6.1926     5.4921  -1.128  0.26325
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.65 on 72 degrees of freedom
## Multiple R-squared:  0.5506, Adjusted R-squared:  0.4819
## F-statistic: 8.018 on 11 and 72 DF,  p-value: 6.326e-09
```

Test for normality



It appears that our data is normally distributed, since the points lie approximately on the center diagonal.

Fisher's test

To determine whether either presence or absence produces greater variability in the winpercent.
I'll check the p.value for the variance and we'll use it to set t.test or Welch's t-test, which adjusts the number of degrees of freedom when the variances are thought not to be equal to each other.

Method: two sided F test to compare two variances

estimate	num df	denom df	p.value	conf.low	conf.high	testedAttribute
0.637	47	36	0.145	0.337	1.171	winpercent by chocolate
2.404	46	37	0.007	1.279	4.426	winpercent by fruity
0.760	70	13	0.450	0.281	1.593	winpercent by caramel
0.622	70	13	0.207	0.230	1.303	winpercent by peanutyalmondy
1.108	77	6	0.995	0.225	2.856	winpercent by nougat
1.763	77	6	0.490	0.357	4.543	winpercent by crispedricewafer
4.500	69	14	0.003	1.732	9.281	winpercent by hard
1.578	63	20	0.257	0.712	3.050	winpercent by bar
1.348	40	43	0.337	0.730	2.508	winpercent by pluribus

If the p.value in F-test is greater than the significance level 0.05 means that there is no significant difference b/n Presence and Absence of an attribute, we reject the Null Hypothesis The higher the deviation of the estimate from 1, the stronger the evidence for unequal variances. 1 is in our confidence interval = variances are equal.

T.test

estimate1	estimate2	statistic	p.value	parameter	conf.low	conf.high	method	alternative	testedAttribute
42.142	60.922	-7.519	0.000	83.000	-23.747	-13.812	Two Sample t-test	two.sided	winpercent by chocolate
42.142	60.922	-7.519	1.000	83.000	-22.934	Inf	Two Sample t-test	greater	winpercent by chocolate
55.327	44.120	3.923	0.000	79.384	5.522	16.893	Welch Two Sample t-test	two.sided	winpercent by fruity
55.327	44.120	3.923	0.000	79.384	6.453	Inf	Welch Two Sample t-test	greater	winpercent by fruity
49.443	60.052	-1.854	0.067	83.000	-21.992	0.775	Two Sample t-test	two.sided	winpercent by nougat
49.443	60.052	-1.854	0.966	83.000	-20.129	Inf	Two Sample t-test	greater	winpercent by nougat
52.418	40.509	4.624	0.000	45.444	6.724	17.095	Welch Two Sample t-test	two.sided	winpercent by hard
52.418	40.509	4.624	0.000	45.444	7.585	Inf	Welch Two Sample t-test	greater	winpercent by hard
46.714	61.295	-4.338	0.000	83.000	-21.266	-7.896	Two Sample t-test	two.sided	winpercent by bar
46.714	61.295	-4.338	1.000	83.000	-20.172	Inf	Two Sample t-test	greater	winpercent by bar
54.066	46.823	2.327	0.022	83.000	1.052	13.436	Two Sample t-test	two.sided	winpercent by pluribus
54.066	46.823	2.327	0.011	83.000	2.065	Inf	Two Sample t-test	greater	winpercent by pluribus
48.894	66.170	-3.127	0.002	83.000	-28.264	-6.289	Two Sample t-test	two.sided	winpercent by crispedricewafer
48.894	66.170	-3.127	0.999	83.000	-26.465	Inf	Two Sample t-test	greater	winpercent by crispedricewafer
47.678	63.697	-4.050	0.000	83.000	-23.886	-8.151	Two Sample t-test	two.sided	winpercent by peanutyalmondy
47.678	63.697	-4.050	1.000	83.000	-22.598	Inf	Two Sample t-test	greater	winpercent by peanutyalmondy
48.931	57.347	-1.990	0.050	83.000	-16.828	-0.005	Two Sample t-test	two.sided	winpercent by caramel
48.931	57.347	-1.990	0.975	83.000	-15.451	Inf	Two Sample t-test	greater	winpercent by caramel

The table shows us for two sided t.test (whether the mean difference is equal to 0) and we see if the p-value is less than 0.05 and we reject the Null Hypothesis. This means that we should include these variable in the model as they have a significant relationship with the winpercent. We would interpret the second t.test (whether the mean difference is greather than 0) the Presence of an attribute in a candy increases winpercent.

Multiple regression finetuning

Based on the F and T test I removed nougat, hard, pluribus, fruity, bar. I also experimented with sugerpercent and pricepercent by making them binary variables.

```
##
## Call:
## lm(formula = winpercent ~ ., data = candy[, -c(1)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.629  -7.137   2.054   5.880  19.270
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)            32.041      3.432   9.337 1.06e-11 ***
## chocolateTRUE          21.280      3.874   5.493 2.27e-06 ***
## caramelTRUE             2.803      3.861   0.726   0.4719
## peanutyalmondyTRUE      9.275      3.553   2.611   0.0126 *
## crispedricewaferTRUE    7.304      4.584   1.593   0.1187
## sugarTRUE               5.282      3.544   1.490   0.1438
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.6 on 41 degrees of freedom
## Multiple R-squared:  0.6042, Adjusted R-squared:  0.556
## F-statistic: 12.52 on 5 and 41 DF,  p-value: 2.098e-07
```

Attribute in candy	Coefficient
(Intercept)	32.041405
chocolateTRUE	21.280251
peanutyalmondyTRUE	9.274791
crispedricewaferTRUE	7.303931
sugarTRUE	5.282090
caramelTRUE	2.803168

How to interpret the model
the presence of an X attribute increases winpercent by its beta Highly significant attributes are the chocolate and peanutyalmondy

The Residuals are symmetrically distributed across 2. That means that the model predicts certain points that fall far away from the actually observed points. Plotting the residuals might be the next step.

Residual Standard Error is a measure of the quality of a linear regression fit. The RSE is the average amount that the winpercent will deviate from the true regression line. In this case on average, it could deviate from the regression line by approximately 11%. The percentage error, in this case, is 33% = any prediction could be off by 33%.

Multiple R-squared, Adjusted R-squared - how well the model is fitting the actual data - the higher the better. We started at 0.47 and ended up at 0.556. = This model fits better the data. hight a regression that does not explain the variance in the response variable well 55,6% of the variance in winpercent can be explained by the predictor variables.

Empfehlung für die Eigenschaften einer neuen Süßigkeit

chocolate, peanutyalmondy, crispedricewafer, caramel, sugar

strictly speaking: having these attributes in a candy would result in a high Win rank.