

**Study on Cereals' Nutrition and Popularity:
Is High Sugar Content Associated With Popularity?**

Introduction

A bowl of cereal is a common breakfast choice for approximately 90% of U.S. households. Even with its 7% drop in volume of sales, cereal is still the most popular breakfast food in the U.S., with an industry profit of over \$9 billion.¹

Several studies have been published on the matter of cereal nutrition and sales. For example, a study published in the Specialty Food Magazine claims that higher sugar content is associated with higher sales.² Interestingly, however, a survey from the same study shows that customers also care about levels of fiber and whole grain content. 52% of a sample of 1544 internet users replied that one of the reasons they eat cereal was for the fiber and grains. Data collection performed by Nielsen Label Trends supports this claim as well.³ Their study shows that the three largest segments by claim are whole grains, vitamin/mineral presence, and fiber presence.

The result that high sugar level is associated with higher sales suggests that, contrary to how several cereal brands are advertised as healthy, the most popular cereal brands may contain high sugar levels that potentially cause health problems. In this study, we will investigate what

¹Information taken from <http://www.cnbc.com/id/100983729>

²Information taken from <http://www.specialtyfood.com/news-trends/featured-articles/article/cereals-reaching-beyond-breakfast/>

³Information taken from <http://www.factsfiguresfuture.com/issues/august-2011/breakfast-cereals-slow-in-supermarkets.html>

factors might influence the popularity of cereals. The primary goal of our analysis is to find out whether sugar content, after controlling other nutritional and environmental factors, is associated with the popularity of cereal. We collected our cereal data from five major cereal companies: Kellogg's, General Mills, Malt-O-Meal, Post, and Quaker Oats. Since our study is a retrospective study, the analysis was done as a test of association involving odds ratio.

Methods

Our data contains 79 cereals from five major cereal manufacturers, with 10 explanatory variables and a response variable indicating the popularity of the cereal. Data for explanatory variables were gathered from the official websites of the cereal manufacturers: Kellogg's, Malt-O-Meal, General Mill, Quaker Oats, and Post. For each manufacturer, we looked at the list of cold breakfast cereal products. Many cereal brands had several varieties within the product, such as Kellogg's Special K, which had 12 varieties just within its Special K brand. In such cases, we picked out two cereals that we deemed most popular (or standard) to avoid double counting cereals that had very similar characteristics. As a result, we constructed an initial data set with 20 cereals from Kellogg's, 15 from General Mills, 17 from Malt-O-Meals, 15 from Post, and 12 from Quaker Oats.

For each cereal listed in our data set, we gathered the calories, sugar content, fiber content, and protein content per 100g to account for the nutritional variables. The nutritional values were collected from the nutrition information sections in the official websites of the five manufacturers.

Our categorical explanatory variables are Name Type (which was categorized as grain, fruit, sweet, or other, judging from the title of a cereal)⁴, crunchy (marked 1 if the name contains some variation of the word, “crunchy,” and 0 otherwise), brand (name of manufacturer), nutritionAds (whether it advertised the presence of certain nutrients on its cereal package), character (whether the cereal package had a cartoon character), and flavors (whether the cereal had other varieties). To collect information concerning the cereal packages, we used photos from each manufacturer’s website, under the assumption that websites used the most standard packaging.

We set values for NameType and Crunchy by simply looking at the titles of the cereal. (For more details, see footnotes¹). We also included information about the advertising strategies used on the cereal boxes in variables, NutritionAd and Character. We used photos from each manufacturer’s website to collect this data, assuming websites to use the most standard packaging. However, we recognize that packaging may vary according to time and place, and that this variable may not be useful for other studies.

Our response variable, which indicates the popularity of the cereal, is a binary variable for whether the cereal is a bestseller(1) or not(0). We set these values after seeing whether it was mentioned in the top ten of at least one of our five arbitrarily picked cereal rankings. The rankings were all determined either by gross sales or by online-surveying. Because the five rankings had substantial overlaps, only 17 cereals were considered bestsellers.

⁴ How we categorized cereals into Name Type groups:

⁴Type Grain: All cereals with some indicator of nuts and grains, such as Kellogg’s Crunchy Nut.

⁴Type Fruit: All cereals with some indicator of fruits, such as Malt-O-Meal’s Apply Zings. We included some whose names implied fruitiness but did not contain actual fruits, such as Froot Loops.

⁴Type Sweet: All cereals with some indicator of sweetness, including those with “honey,” “sugar,” and “chocolate” in their names.

⁴In case that the title included more than one type, we considered the word that comes first. For example, Raisin Bran is both a fruit and grain type, but we considered it as Type Fruit.

Results

Adjusting the Data Set and Variable Overview

After examining our dataset, we removed cereals, Puffed Rice and Puffed Wheat, from our analysis, because they both contained one missing value each. We decided not to log-transform any of our variables, either because log-transforming did not help the variable's linear relationship with the logit of the TopTen variable, or because we wanted to check for the variable's curvature.

An explanatory data analysis was performed with all of our variables: NameType, Crunchy, Manufacturer, Calories, Sugar, Fiber, Protein, NutritionAds, Character, and Flavors. Below are visual representation of our data distribution for each variable.⁵ Each visual representation was created without cereal Manufacturers, Puffed Rice (Quaker Oats), Puffed Wheat (Quaker Oats), and Special K (Kellogg's). Special K was flagged as an outlier during the process of fitting a logistic regression model to our data.

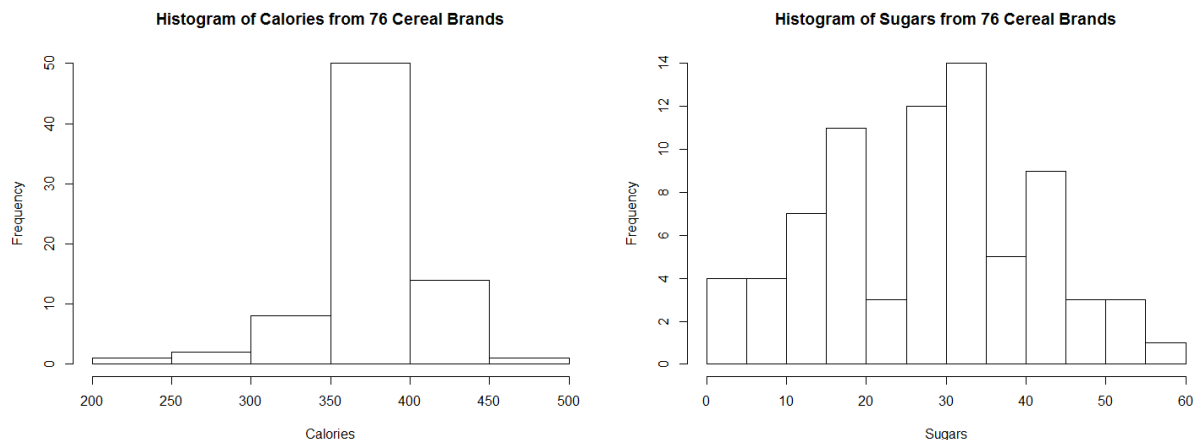


Figure 1 and 2. Empirical distributions for Calories (in calories) and Sugars (in grams)

⁵ Exploratory data analysis other nutritional variables are in the Appendix C section.

	Min	1st Quartile	Median	Mean	3rd Quartile	Max	Std. Deviation
Calories	200.00	368.00	387.10	378.70	400.00	476.20	39.04
Sugars	0.00	18.23	29.06	28.05	37.26	55.56	13.45
Fibers	0.00	3.31	6.19	7.38	10.21	46.67	7.62
Protein	3.13	3.70	6.67	6.86	8.59	14.55	2.894

Table 1. Summary statistics for Calories, Sugars, Fibers, and Protein for 76 cereal brands

	Character		Flavors		NutritionAds	
	Yes	No	Yes	No	Yes	No
Top 10: Yes	50	10	42	18	30	30
Top 10: No	3	13	5	11	3	13

Table 2. Placement in Top 10 Rankings Against Character, Flavors, and NutritionAds

Assumptions

One assumption we make about our data is that the sugar and fiber values for Puffed Rice and Puffed Wheat (both from Quaker Oats) were missing at random. Because of this assumption, we were able to simply remove the two cereal brands from our data set. For the sake of this assignment, we also assume that our observations are independent and that our independent explanatory variables are linearly related to the logit of our TopTen variable.

Fitting a Logistic Regression Model

We started our model building by fitting a rich logistic regression model that included all 10 variables in our data, except for Sugars. We added variable Sugars only after we obtained the final model, because we wanted to know the association between sugar and top-ten, after controlling the other variables. In the early stages of model-building, we dropped variable Manufacturer, because none of the cereals in our Malt-O-Meal sample was marked as being in at least one of 5 top-ten rankings.

	General Mills	Kellogg's	Malt-O-Meals	Post	Quaker Oats
Top 10: No	8	13	17	13	9
Top 10: Yes	7	6	0	2	1

Table 3. Counts of Cereals in Rankings for Each Manufacturer

After further model building process, we found that only variables Calories, NutritionAds, Character and Flavor were significant. We removed Kellogg's Special K from our analysis because it was flagged as an influential outlier during our model building process. We suspect this was because it was the only cereal brand in our data set that did not have any nutritional advertisements nor cartoon characters on its box in the picture used on the manufacturer's official website. Below is the model equation for our final model without sugar:

$E[\text{top10} \mid \text{Calories, NutritionAds, Character, Flavors}]$

$$= \beta_0 + \beta_1(\text{Calories}) + \beta_2(\text{NutritionAds}) + \beta_3(\text{Character}) + \beta_4(\text{Flavors})$$

The estimates of the model parameters are given below. We kept variable Calories in our model although it is not significant at the 0.05 significance level because the extra sum of squares test indicated that it should be kept it.

	Estimate	Std. Error	z-value	p-value
Intercept	-21.92509	8.3602	-2.623	0.00873
Calories	0.03527	0.01839	1.92	0.05513
NutritionAds	3.13018	1.42641	2.194	0.0282
Character	4.64723	1.42556	3.26	0.00111
Flavors	3.48789	1.31403	2.654	0.00795

Table 4. Model Estimates for Adjusting Variables

Here are our interpretations for this model. The categorical variables are interpreted under the assumption that all other variables are kept constant.

There is not enough evidence to conclude that calories is significant because the 95% confidence interval for odds includes 1 (95% CI: 0.999 to 1.074). However, the odds (of being in top-ten) for a cereal brand that advertises its nutrition content on its box are approximately 22.878 times higher than the odds of a cereal brand that does not. (With 95% confidence, the odds are 1.397 to 374.632 times higher.) The presence of a cartoon character on the box is significant as well, with the odds of a cereal brand with a cartoon character on a top-ten list being 104.296 times higher than the odds of one without. (With 95% confidence, the odds are 6.379 to 1705.012 times higher.) Lastly, the odds of a cereal brand with other flavors being on a top-ten list are 32.717 times higher than the odds of a cereal brand without other flavors. (With 95% confidence, the odds are 2.490 to 429.830 times higher.)

We recognize that the 95% confidence range is very wide, and we attribute this to the fact that the standard error for each variable is large. We checked for multicollinearity and found that each variable had a multicollinearity level of under 2.5. Therefore, we suspect that the wide confidence interval is either because 1) we have a lot of categorical variables of yes/no or because 2) the ranges of our numerical variables are small (For example, the min-max range of the fiber variable is 0-6).

Accounting for Sugars

After accounting for the effects of calories, the existence of nutrition-related claims and cartoon characters on boxes, and whether or not the cereal brand had other flavors or variations, we included variable Sugars into our model. Let us consider our final model:

$$E[\text{top10} \mid \text{Calories, NutritionAds, Character, Flavors, Sugars}]$$

$$= \beta_0 + \beta_1(\text{Calories}) + \beta_2(\text{NutritionAds}) + \beta_3(\text{Character}) + \beta_4(\text{Flavors}) + \beta_5(\text{Sugars})$$

The estimates for our Sugars parameter are given below.

	Estimate	Std. Error	z-value	p-value
Sugars	0.02539	0.04167	0.609	0.5423

Table 5: Model Estimates for Sugars

We do not have enough evidence to conclude that there is a difference in odds for a cereal brand with $x+1$ grams of sugar compared to the odds for one with x grams of sugars. (Our 95% confidence interval for odds difference is 0.945 to 1.113).

DISCUSSION

Our dataset consists of 76 cereals with 10 exploratory variables and a binary response of whether the cereal was mentioned in one of our top-ten lists or not. We fit a binary logistic regression model to examine which factors are associated with high sales and to find out whether sugars content in the cereal is associated with the TopTen variable, after controlling for all other variables. Our result indicates that high levels of sugar are not associated with a cereal's popularity. Instead, we found that caloric values, presence of nutritional claims and cartoon characters, and variations of the brand are associated with a cereal brand's popularity. However, our model did not give us very clean interpretations of the data. We suspect that one reason is

that there were great overlaps in our numeric value. For example, despite our suspicions, the median level of sugar did not differ between cereal brands that were in top-ten rankings and cereal brands that weren't.

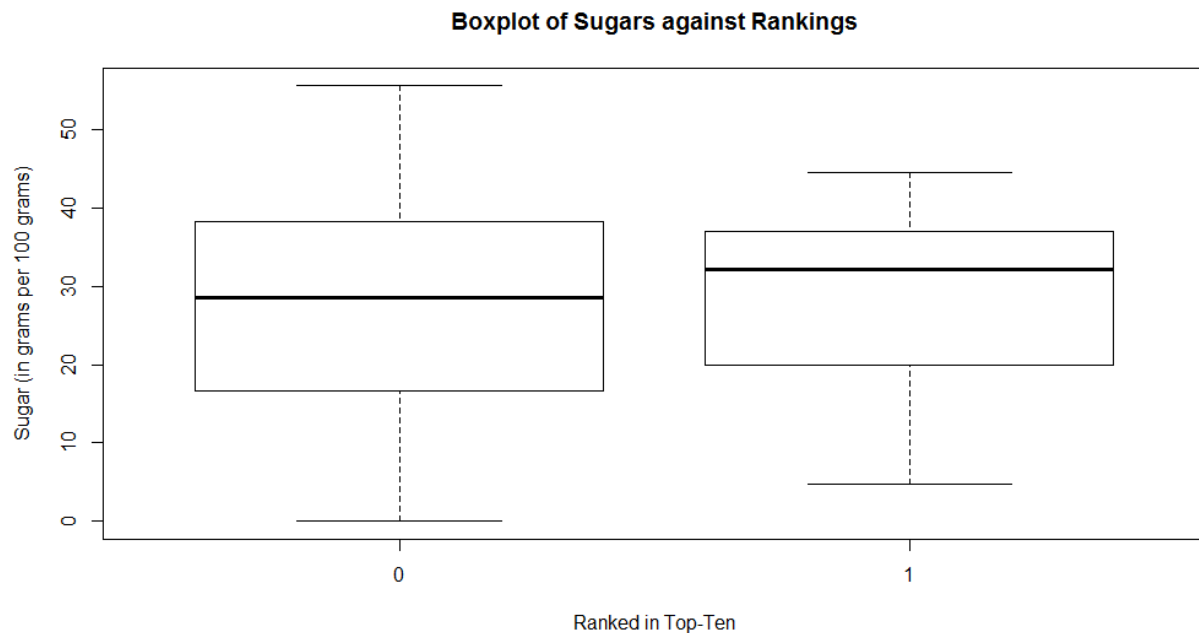


Figure 6: Boxplot of Sugars Against Rankings

Another limitation of our study is caused by the methods we used for our data collection. We used a convenience sample by arbitrarily picking one or two standard or most popular variations for each cold-cereal brand from five manufacturers. The TopTen variable was chosen with information collected from online survey websites--Lava Surfer, Food Navigator, Rankings.com, Top Ten, and Complex--, which further limits the accuracy of our study because we could not figure out exact sources of the rankings and therefore could not account for any correlated data among the rankings.

We made many assumptions about our dataset in order to use the logistic regression model. One of our biggest assumptions was that our cereal observations were independent. This

may not be the case because manufacturers may change ingredients and produce new brands based on what kinds of cereals sell the most. For example, if a cereal brand with a colorful cartoon character is often ranked high on cereal rankings, manufacturers will probably use this information to sell more cereals by attaching characters to brands that did not previously have characters.

We also recognize that there is some degree of vagueness in our definitions for the categorical variables: NutritionAds and Character. Because we categorized each cereal based on our own personal judgements and because the information came from official website photos which are prone to frequent change, the data is not applicable to other studies. We must also recognize that both NameType and Crunchy may not accurately represent our data. For example, we put chocolate-type cereals and honey-type cereals in the same category, but consumer groups who prefer chocolate-type cereals may have different buying patterns compared to consumer groups that prefer the honey-type. Concerning variable Crunchy, some consumers may judge a cereal brand's crunchiness based on visuals on the box instead of the name, as we did.

For future studies, we would like to suggest investigating the same question using a numeric variable indicating the profit for each product. A multiple regressions model may give cleaner results because it would account for more detail. We would also like to suggest using variables that can be used repeated for other studies because we suspect that the personal bias in our categorical variables affected the accuracy of this model.

Lastly, we would like to use the results of this model to suggest that a more extensive study be done on advertising strategies and sales for cereals. Three out of the four significant variables in our final model (NutritionAds, Character, and Flavor) were related to how a cereal product was advertised or presented. This strongly suggests that there may be a significant

association between big cereal manufacturers and how much they spend on advertisements or the popularity of a cereal brand and how unique its advertising strategy is.

Appendix

Appendix A

Name	NameType	Crunchy	Brand	Calories	Sugars	Fibers	Protein	NutritionAc	Character	Flavors	Response
Frosted Flakes	Other	1	Kellogg's	379.3103448	34.48275862	0	3.448275862	1	1	1	1
Froot Loops	Fruit	0	Kellogg's	379.3103448	41.37931034	10.34482759	3.448275862	1	1	0	1
Crunchy Nut	Grain	1	Kellogg's	387.0967742	32.25806452	0	6.451612903	0	0	0	0
Mini-Wheats	Grain	0	Kellogg's	351.8518519	20.37037037	11.11111111	9.259259259	1	1	1	1
Krave (Chocolate)	Sweet	0	Kellogg's	363.6363636	33.33333333	9.090909091	6.060606061	1	0	1	0
Rice Krispies (Original)	Grain	1	Kellogg's	393.9393939	12.12121212	0	6.060606061	1	1	1	1
Special K with Red Berri	Fruit	0	Kellogg's	354.8387097	29.03225806	9.677419355	6.451612903	1	0	1	0
Original Special K	Other	0	Kellogg's	387.0967742	12.90322581	0	19.35483871	0	0	1	1
All-Bran (Original)	Grain	0	Kellogg's	258.0645161	19.35483871	32.25806452	12.90322581	1	0	1	0

Appendix B

The information used throughout this report was taken from these following sources:

Data for each cereal brand:

http://www.kelloggs.com/en_US/product-search.pt-Cereal*.html

<http://www.generalmills.com/en/Brands/Cereals.aspx>

<http://www.postfoods.com/our-brands/>

http://www.quakeroats.com/products.aspx?view=OatCereal&gclid=CjkKEQjwk9CcBRDEopHmnZa5td8BEiQAr2BckKC8RYg8Wh_6JmmSm09gAk9m4lR3DCMGxABH3hgciKrw_wcB

<http://www.malt-o-meal.com/cold-cereals/>

Data for top 10 rankings:

<http://www.lavasurfer.com/cereal-stats.html>

[http://www.foodnavigator-usa.com/Manufacturers/Cereal-blockbusters-America-s-top-10-best-selling-brands/\(page\)/1](http://www.foodnavigator-usa.com/Manufacturers/Cereal-blockbusters-America-s-top-10-best-selling-brands/(page)/1)

<http://www.ranker.com/crowdranked-list/the-best-breakfast-cereals-of-all-time>

<http://www.thetoptens.com/cereals/>

<http://www.complex.com/city-guide/2013/01/the-50-greatest-breakfast-cereals-of-all-time/>

Source	Lava Surfer (top 15)	Food Navigator	Ranker (top 20)	The Top Ten	Complex (top 50)
1	Cheerios	Honey Nut Cheerios	Frosted Flakes	Cinnamon Toast Crunch	Cheerios
2	Special K	Frosted Flakes	Captain Crunch	Captain Crunch	Cinnamon Toast Crunch
3	Frosted Flakes	Honey Bunches of Oats	Lucky Charms	Lucky Charms	Wheaties
4	Honey Bunches of Oats	Cheerios	Cinnamon Toast Crunch	Frosted Flakes	Mini Wheats
5	Frosted Mini Wheats	Cinnamon Toast Crunch	Honey Nut Cheerios	Cheerios	Honey Nut Cheerios
6	Cinnamon Toast Crunch	Special K	Froot Loops	Fruity Pebbles	Apple Jacks
7	Lucky Charms	Frosted Mini Wheats	Apply Jacks	Reese's Puffs	Rice Krispies
8	Captain Crunch	Lucky Charms	Captain Crunch (Berries)	Apple Jacks	Lucky Charms
9	Raisin Bran	Froot Loops	Cocoa Puffs	Coco Pops	Raisin Bran
10	Froot Loops	Raisin Bran	Rice Krispies	Frosted Mini Wheats	Count Chocula

Appendix C

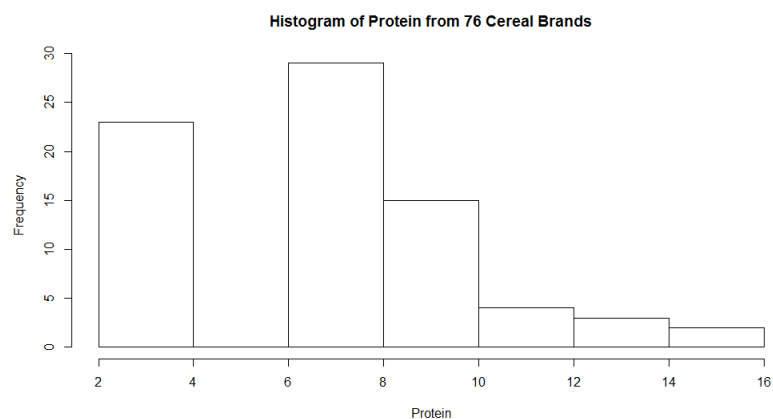


Figure 1: Histogram of Protein from 76 Cereal Brands

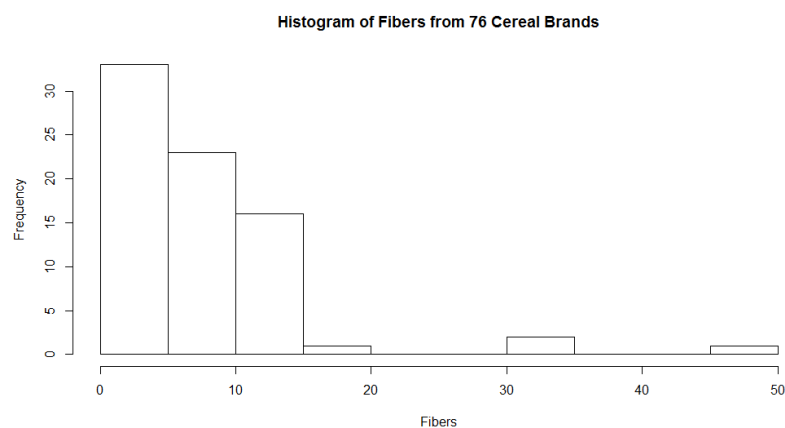


Figure 2: Histogram of Fibers from 76 Cereal Brands

Appendix D

Final Project R Script with CerealData – Risako Owan and Jimin Yoo-

```
head(CerealData)
```

Adjusting data set: figure out NAs

```
library(Hmisc)
na.patterns <- naclus(CerealData)
naplot(na.patterns, which = 'na per var')
CerealData$missSugar <- ifelse(is.na(CerealData$Sugars), 1, 0)
which(CerealData$missSugar == 1) #43
CerealData$missFibers <- ifelse(is.na(CerealData$Fibers), 1, 0)
which(CerealData$missFibers == 1) #42

#Take out 42,43
CerealData2 <- CerealData[-c(42,43),]
```

Exploratory Data Analysis

```
summary(CerealData2$Calories)
sd(CerealData2$Calories)
summary(CerealData2$Sugar)
sd(CerealData2$Sugar)
summary(CerealData2$Fiber)
sd(CerealData2$Fiber)
summary(CerealData2$Protein)
sd(CerealData2$Protein)

table(CerealData2$TopTen, CerealData2$NameType)
table(CerealData2$TopTen, CerealData2$Manufacturer)
#Malt-O-Meal has no cereals in the top 10
table(CerealData2$TopTen, CerealData2$Crunchy)
table(CerealData2$NameType, CerealData2$Crunchy)
table(CerealData2$TopTen, CerealData2$NutritionAds)
table(CerealData2$TopTen, CerealData2$Character)
table(CerealData2$TopTen, CerealData2$Variety)
table(CerealData2$TopTen)
dev.off()
boxplot(Sugars~Manufacturer, data=CerealData[-c(42,43),])
boxplot(Fibers~Manufacturer, data=CerealData2)
boxplot(Protein~Manufacturer, data=CerealData2)
boxplot(Calories~Manufacturer, data=CerealData2)

plot(TopTen~Sugars, data=CerealData2, xlab="Sugars", ylab="Probability of Being in Top 10")
title("Top 10 against Sugars")
plot(TopTen~Fibers, data=CerealData2, xlab="Fibers", ylab="Probability of Being in Top 10")
```

```

title("Top 10 against Fibers")
plot(TopTen~Protein, data=CerealData2, xlab="Protein", ylab="Probability of Being in Top 10")
title("Top 10 against Protein")
plot(TopTen~Calories, data=CerealData2, xlab="Calories", ylab="Probability of Being in Top 10")
title("Top 10 against Calories")

```

```

library(car)
scatterplotMatrix(CerealData2[, -c(1,2,11,14,15)], diagonal="histogram", smooth=FALSE,
reg.line=FALSE)

```

```

head(CerealData2)
scatterplotMatrix(CerealData2[, c(13,5,6,7,8)])

```

Model Building Process and Diagnostics

```

cereal.glm <-
glm(TopTen~NameType+Crunchy+Calories+Fibers+I(Fibers^2)+Protein+NutritionAds+Character+Variety+Calories:Fibers, data=CerealData2, family=binomial)
summary(cereal.glm)

```

```

plot(resid(cereal.glm, type="response"))
which(resid(cereal.glm, type="response") > 0.9) #8,62(60)
which(resid(cereal.glm, type="response") < -0.9)
plot(cereal.glm, which=4) #8
which(hatvalues(cereal.glm) > 2*11/77) #a lot

```

#We took out the influential outlier 8, Special K.

```

CerealData3 <- CerealData2[-c(8,60),]
cereal.glm2 <-
glm(TopTen~NameType+Crunchy+Calories+Fibers+I(Fibers^2)+Protein+NutritionAds+Character+Variety+Calories:Fibers, data=CerealData3, family=binomial)

```

```

ftable(xtabs(cbind(TopTen, 1 - TopTen) ~ NameType+Crunchy, data = CerealData2))

```

#None of the crunchy Fruit and sweet cereals was in the top 10, which causes problem in our model. We took out the NameType variable, because it turned out not significant in testing the first model, and also because we suspected the variable is not reliable.

#Test model (not including NameType and Manufacturer) with only 42 and 43 taken out

```

cereal.glm3 <-
glm(TopTen~Crunchy+Calories+Fibers+I(Fibers^2)+Protein+NutritionAds+Character+Variety+Calories:Fibers, data=CerealData2, family=binomial)
summary(cereal.glm3)
#Diagnostics of cereal.glm3
plot(resid(cereal.glm3, type="response"))

```

```

which(resid(cereal.glm3, type="response") > 0.9)#8,62
which(resid(cereal.glm3, type="response") < -0.9)
plot(cereal.glm3, which=4) #8
which(hatvalues(cereal.glm3) > 2*12/77) #a lot

CerealData4 <- CerealData2[-c(8),]
cereal.glm4 <-
glm(TopTen~Crunchy+Calories+Fibers+I(Fibers^2)+Protein+NutritionAds+Character+Variety+Calories:
Fibers, data=CerealData4, family=binomial)
summary(cereal.glm4)
#60 is flagged probably because it is the only best-selling cereal without a cartoon character and Variety.
Leave 60 in because the cook's distance and leverage do not flag it
#Take out 8 because it is flagged both by the cook's distance and by the leverage

#Next, Fiber is taken out, because it turns out insignificant, and the EDA shows there was no difference in
median between the two popularity groups boxplot(Fibers~TopTen, data=CerealData4)

cereal.glm5 <- glm(TopTen~Crunchy+Calories+Protein+NutritionAds+
Character+Variety,
data=CerealData4, family=binomial)
summary(cereal.glm5)
anova(cereal.glm5, cereal.glm4, test="Chisq")#We prefer the smaller model
#Diagnostics of cereal.glm5
plot(resid(cereal.glm5, type="response"))
which(resid(cereal.glm5, type="response") > 0.9)#8,62(60)
which(resid(cereal.glm5, type="response") < -0.9)
plot(cereal.glm5, which=4) #None over 1
which(hatvalues(cereal.glm5) > 2*7/76) #a lot

cereal.glm6 <- update(cereal.glm5, .~-Crunchy)
anova(cereal.glm6, cereal.glm5, test="Chisq")#Yay
summary(cereal.glm6)
#Diagnostics of cereal.glm6
plot(resid(cereal.glm6, type="response"))
which(resid(cereal.glm6, type="response") > 0.9)#8,62(60)
which(resid(cereal.glm6, type="response") < -0.9)
plot(cereal.glm6, which=4) #None over 1
which(hatvalues(cereal.glm6) > 2*6/76) #a lot

cereal.glm7 <- update(cereal.glm6, .~-Protein)
anova(cereal.glm7, cereal.glm6, test="Chisq")#0.07 (insignificant at the 0.05 significance level)
summary(cereal.glm7)
#Diagnostics of cereal.glm7
plot(resid(cereal.glm7, type="response"))

```

```
which(resid(cereal.glm7, type="response") > 0.9)#8,62(60)
which(resid(cereal.glm7, type="response") < -0.9)
plot(cereal.glm7, which=4) #None over 1
which(hatvalues(cereal.glm7) > 2*5/76) #a lot
```

```
cereal.glm8 <- update(cereal.glm7, .~.-Calories)
anova(cereal.glm8, cereal.glm7, test="Chisq")
#Larger model is preferred
#Proceed with cereal.glm7
```

#Now add Sugars

```
cereal.glm9 <- update(cereal.glm7, .~.+Sugars)
anova(cereal.glm9, cereal.glm7, test="Chisq")
summary(cereal.glm9)
plot(resid(cereal.glm9, type="response"))
which(resid(cereal.glm9, type="response") > 0.9)
which(resid(cereal.glm9, type="response") < -0.9)
plot(cereal.glm9, which=4) #None over 1
which(hatvalues(cereal.glm9) > 2*7/76) #a lot
```

#check 17

```
cereal.glm10 <- update(cereal.glm9, .~., subset=-17)
summary(cereal.glm10)
#Leave 17 in, it is not an influential outlier
```

Multicollinearity Check

```
vif(cereal.glm9)
```

Final Model Estimates

```
summary(cereal.glm9)
```