

# FinalProject

June 1, 2023

#

Final Project

```
[2]: # Load the CourseKata library
      suppressPackageStartupMessages({
        library(coursekata)
      })
```

## 0.1 1. Intro/Overview of the Problem or Question

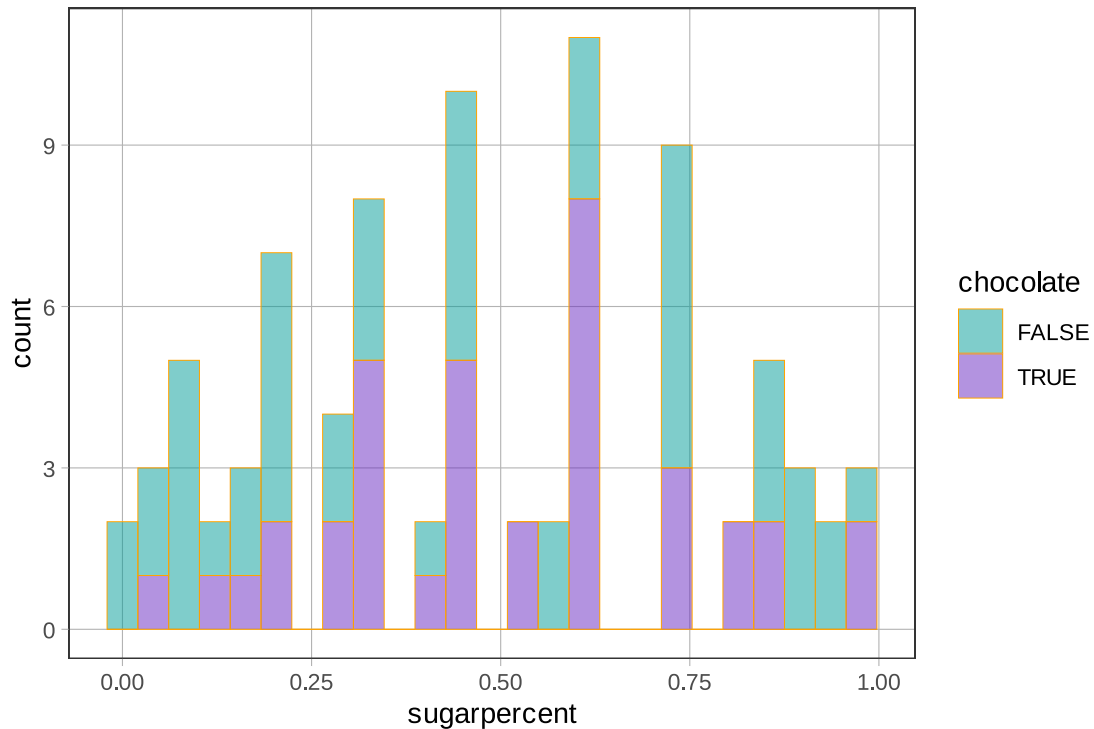
I am wondering if chocolate has the most sugar. This question is important for kids that go trick or treating, so parents can limit their sugar intake. The data comes from 8,371 different IP addresses on ~269,000 randomly generated matchups. This data was collected to provide treats that will satisfy the trick or treaters. The dataset includes the candy names, the kind of candy, sugar percent, price, and the percent of how many people would eat it. I predict that chocolate candies have the most sugar and I think it is possible because chocolate is very sweet.  $\text{sugarpercent} = \text{chocolate} + \text{Other Stuff}$

```
[3]: (candy_rankings)
```

	competitorname <chr>	chocolate <lgl>	fruity <lgl>	caramel <lgl>	peanutyalmondy <lgl>	nougat <lgl>	candy <lgl>
A tibble: 85 × 13	100 Grand	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE
	3 Musketeers	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE
	One dime	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
	One quarter	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
	Air Heads	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Almond Joy	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE
	Baby Ruth	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE
	Boston Baked Beans	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
	Candy Corn	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
	Caramel Apple Pops	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
	Charleston Chew	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE
	Chewey Lemonhead Fruit Mix	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Chiclets	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Dots	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Dum Dums	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Fruit Chews	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Fun Dip	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Gobstopper	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Haribo Gold Bears	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Haribo Happy Cola	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
	Haribo Sour Bears	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Haribo Twin Snakes	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Hershey's Kisses	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
	Hershey's Krackel	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE
	Hershey's Milk Chocolate	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
	Hershey's Special Dark	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
	Jawbusters	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Junior Mints	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
	Kit Kat	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE
	Laffy Taffy	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Ring pop	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Rolo	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE
	Root Beer Barrels	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
	Runts	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Sixlets	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
	Skittles original	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Skittles wildberry	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Nestle Smarties	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
	Smarties candy	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Snickers	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE
	Snickers Crisper	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE
	Sour Patch Kids	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Sour Patch Tricksters	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Starburst	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Strawberry bon bons	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Sugar Babies	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
	Sugar Daddy	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
	Super Bubble	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Swedish Fish	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
	Tootsie Pop	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE

## 0.2 2. Explore Variation

```
[2]: gf_histogram(~sugarpercent, data = candy_rankings, fill = ~chocolate, color = "orange")
```



There is no missing data and I did not need to create new variables. The candies with chocolate have the most sugar. While candies without chocolate have the least sugar. The histogram shows the sugar percentage side by side. My hypothesis of chocolate having more sugar is proven in this histogram.

## 0.3 3. Model Variation

```
[6]: lm(chocolate~sugarpercent, data = candy_rankings)
chocolate.model <- lm(chocolate~sugarpercent, data = candy_rankings)
```

Call:

```
lm(formula = chocolate ~ sugarpercent, data = candy_rankings)
```

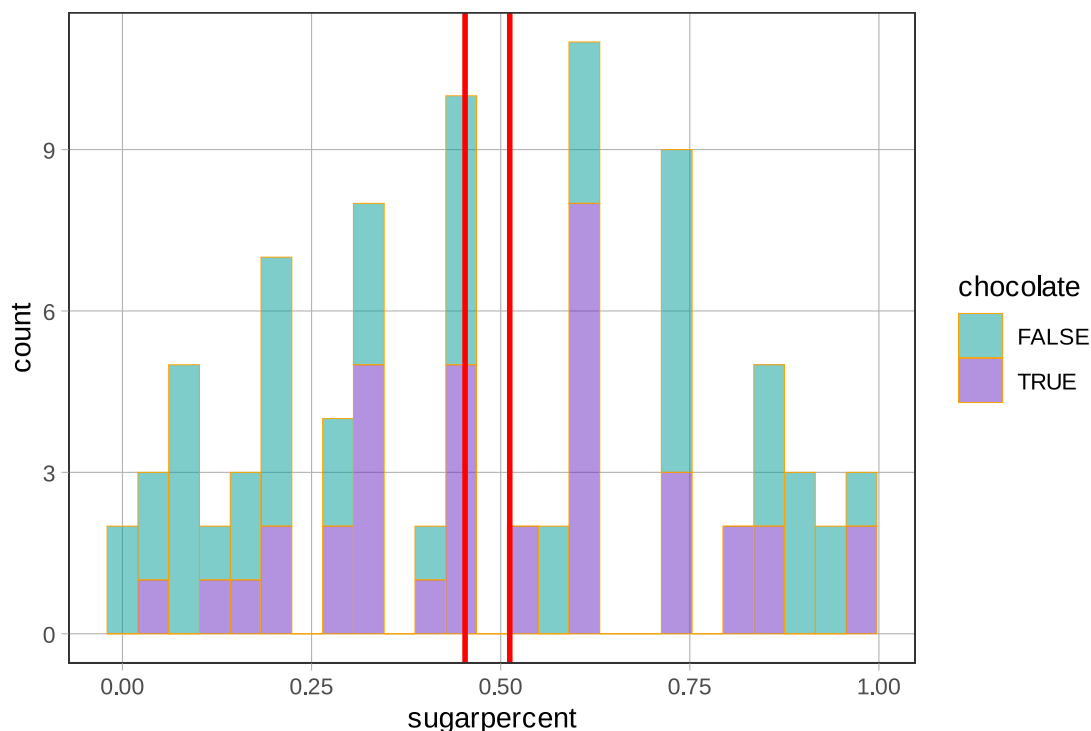
Coefficients:

```
(Intercept)  sugarpercent
      0.3474      0.1837
```

$= 0.3474 - .1837 + e_i$

The mean for the amount of sugar in chocolate is 0.3474. The difference of the mean prediction between chocolate and other candies are 0.1837. The empty model shows that chocolate has more sugar than other candies.

```
[5]: gf_histogram(~sugarpercent, data = candy_rankings, fill = ~chocolate, color = "orange") %>%
      gf_model(sugarpercent ~ chocolate, data = candy_rankings, color = "red")
```



The red line indicates that 0.50 is the mean of the amount of sugar in the candies.

## 0.4 4. Evaluate Models

```
[4]: supernova(chocolate.model)
```

Analysis of Variance Table (Type III SS)

Model: chocolate ~ sugarpercent

	SS	df	MS	F	PRE	p
Model (error reduced)	0.227	1	0.227	0.911	.0109	.3427
Error (from model)	20.667	83	0.249			
Total (empty model)	20.894	84	0.249			

The sum of square residuals is 20.894 and 20.667 is the sum of residuals for the group model. The

chocolate model explained 0.227 out of 20.667 of the total error. The PRE is .0109, which is the amount of error reduced.

```
[9]: PRE(chocolate.model)
```

```
0.0108511935114047
```

## 0.5 5. Conclusions

I examined if chocolate has more sugar than other candies. I used a histogram to show the amount of sugar in chocolate and other candies. The chocolate model proved that they have more sugar. It relates to the question because it will help parents limit the sugar intake for kids. The mean for sugar in chocolate is 0.3474 is higher than 0.1837, which is lower than the mean for sugar in other candies.