Bryan Greener

Chad Hirsch

Jason Gunderson

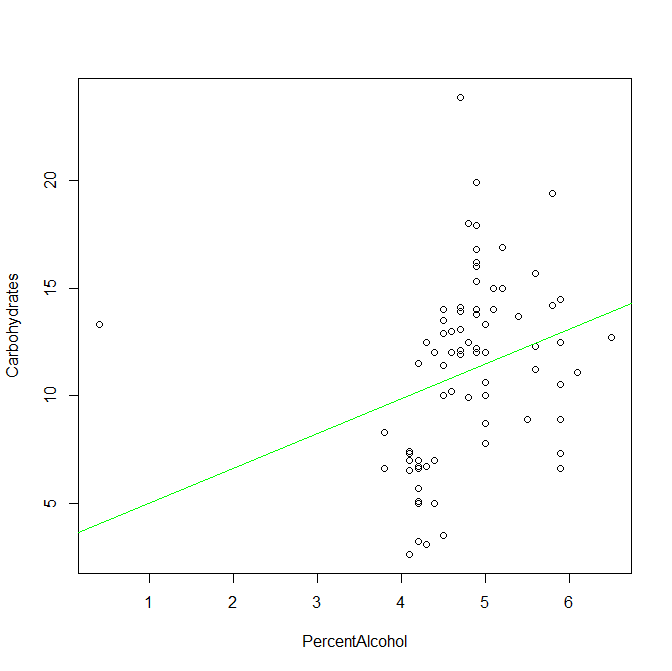
2-3 Workshop 2017-01-30

2.71



Beer with Outlier and Regression Line

Yhat=11.06744-1.615464\*4.759302

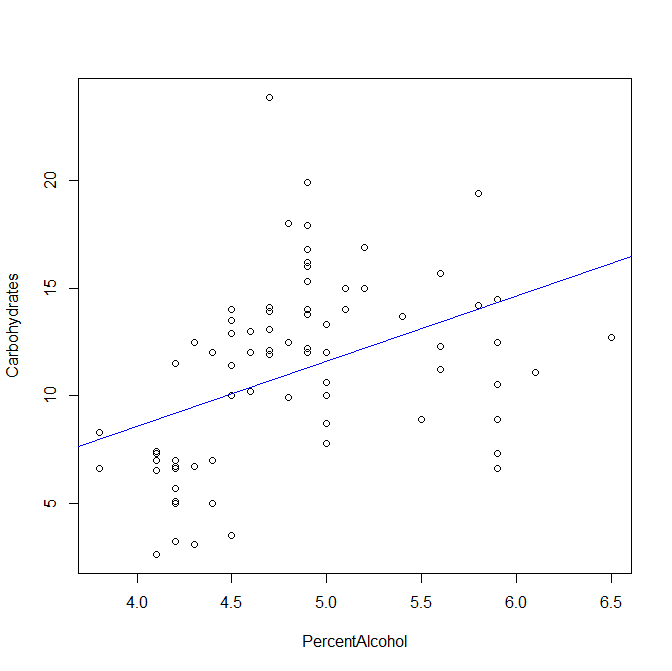


2.72



Beer without Outlier and with Regression Line

Yhat=11.04118-3.031908\*4.810588



The outlier in this data caused the slope to be much lower than when the outlier is removed. With the outlier still in the dataset, the regression line does fit that set of data better but for the bulk of the values the regression line is thrown off. After removing the outlier, the regression line fits the bulk of the points better making the strength of the line higher.

Code below:

#read file

beers = read.csv("beer.csv")

attach(beers)

names(beers)

# set variables for the reg line formula

xbar=mean(PercentAlcohol)

xbar

ybar=mean(Carbohydrates)

ybar

sx=sd(PercentAlcohol)

sx

sy=sd(Carbohydrates)

sy

r=cor(PercentAlcohol,Carbohydrates)

r

#calculate regression line

slope=r\*sy/sx

intercept=ybar-slope\*xbar

#yhat=ybar-slope\*xbar

#plot data and display reg line

plot(PercentAlcohol,Carbohydrates)

fit=lm(Carbohydrates ~ PercentAlcohol)

abline(fit,col='green')

#remove outlier

which(PercentAlcohol==min(PercentAlcohol))

PercentAlcohol[57]

beerOutlier=beers[-57,]

detach(beers)

#attach new data set

attach(beerOutlier)

plot(PercentAlcohol,Carbohydrates)

cor(PercentAlcohol,Carbohydrates)

#set variables for reg line calc

xbar=mean(PercentAlcohol)

xbar

ybar=mean(Carbohydrates)

ybar

sx=sd(PercentAlcohol)

sx

sy=sd(Carbohydrates)

sy

r=cor(PercentAlcohol,Carbohydrates)

r

#calculate regression line

slope=r\*sy/sx

intercept=ybar-slope\*xbar

#add reg line to new plot

fit=lm(Carbohydrates ~ PercentAlcohol)

abline(fit,col='blue')

detach(beerOutlier)