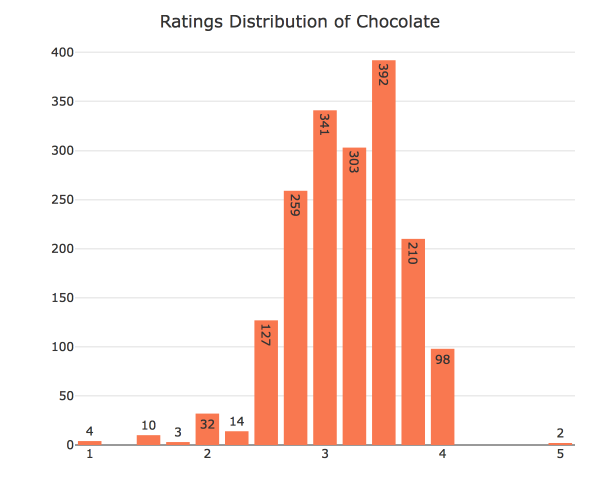
Chocolate Bar Ratings

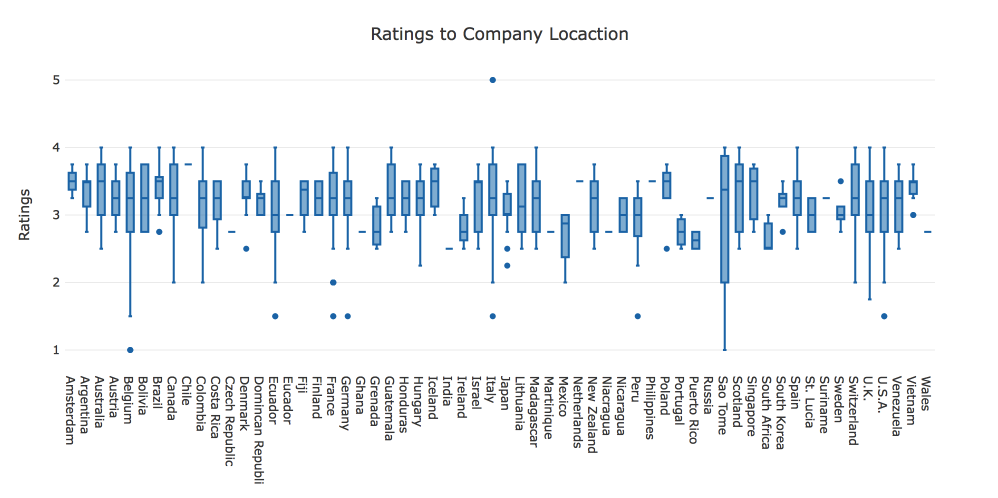
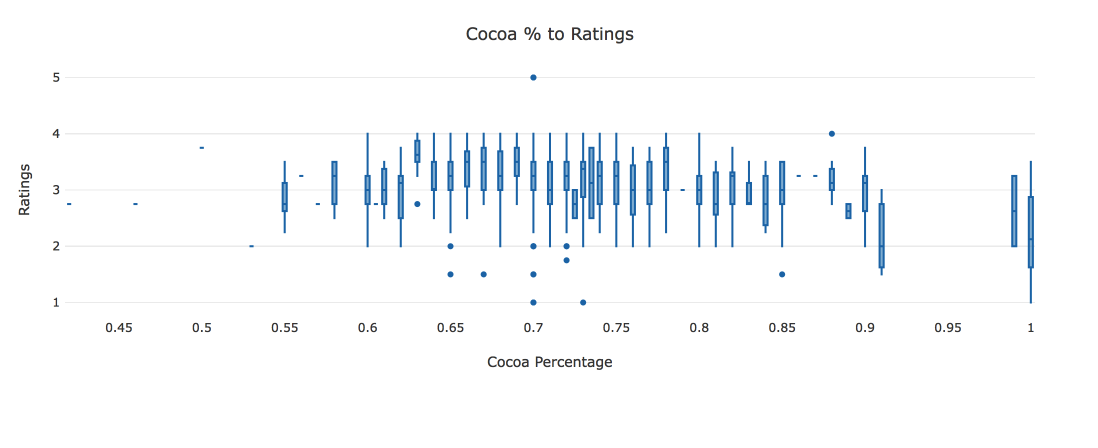
by Ali Awari

I’m currently a Junior attempting a Bachelors of Art in Computer Science. My intention is to become a full-stack web developer, with an aptitude for kernel development.

For this week’s assignment, I analyzed a data set on chocolate ratings, which I found from: <https://www.kaggle.com/rtatman/chocolate-bar-ratings>. I thought it would be interesting to analyze this dataset because, chocolate is an important part of US culture. It’s in protein shakes, bars, and desserts. In this context the chocolate is in the form of a bar.

First thing I looked at when analyzing this dataset is the distribution of ratings to get a feeling of where good chocolate and bad chocolate lies among the reviewers. Good chocolate resides around a rating of *4* and bad chocolate resides around a *2.5* rating. Now to actually see what this dataset holds.

The dataset offers information about the location of the chocolate companies producing the chocolate bars that have been reviewed. It tends to be said that “Belgium chocolates are the best” or “Germans have a superior chocolate”. Would it be interesting to see if that is true? I’d think so, R allows us to do this by graphing all of the countries in terms of mean rating.

 The data shows that companies with the highest ratings are located in: Chile, Amsterdam, Netherlands, Philippines, and Iceland. I think I would sound redundant attempting to find out the best place to buy chocolates from, instead I will find the places with the worst chocolate. •Sao Tome, Peru, Grenada, Ireland, Czech Republic, Ghana, Martinique, Nicaragua, Portugal, Wales, Mexico, South Africa, Puerto Rico and India are the nations with average ratings under 3.

As it turns out higher levels of cocoa also seems to be inversely connected with ratings, as higher percentages of cocoa tend to have lower ratings. Sao Tome, Ghana, Wales, Ireland, Portugal and Nicaragua have an average of 70% cocoa in their chocolates. Assuming that lower cocoa might just be personal taste on the reviewer, I came up with a hypothesis that chocolates from Nicaragua are worse than most other chocolates. To prove that my data is not random I choose to do a Z-test in order to find out. My zeta ended up being 1.659215, and my P-value ended up at 0.04853625 which allows me to deny a null hypothesis. Clearly speaking there might be other variables that this dataset had not accounted for however, I am happy to say that the data is not random. Thank you for reading my post.

* Ali Awari