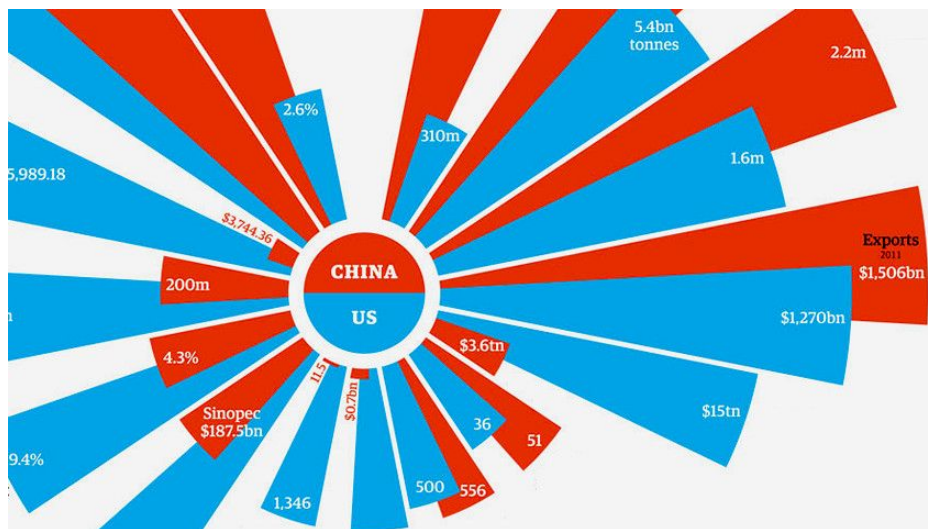


We will use the Amazon data set from <http://jmcauley.ucsd.edu/data/amazon/>

We are going to visualize the number of 5-star ratings different Amazon retail categories get. The abstract question we are trying to ask is “which amazon category are you most likely to be happy purchasing a product from?” We’ve found a dataset of over 100 million reviews linked to different categories. We haven’t decided if we are going to process all the data or only a subset of it. Either way, we plan on using python to extract data points such as product category and rating.

We are currently between two strategies for visualizing the Amazon data set. One of our ideas is a rose petal visualization. If we use this, we would have each product category as its own dx of the circle and each dx would have its own bar graph based on the ratings.



The other strategy that we are considering is a bubble chart similar to the one below but with modifications. We want to either divide each bubble in 5 sections with each section belonging to a star rating or vary the hues based on whichever rating is top heavy in each category. Finally, we will scale the bubble based on the percentage that the 5-star rating takes up on the bubble.

