-- (10 pts) Q1: What are the top 20 cities (with state information) with the most businesses on yelp?

-- Expected columns: state, city, business\_count

-- (10 pts) Q2: What are the average stars for top 10 most reviewed restaurants in Toronto?

-- Expected columns: business\_id, review\_count, avg\_star

-- (10 pts) Q3: How many users have left reviews but all the reviews are 1 star?

-- Expected columns: user\_count

-- (15 pts) Q4: Who wrote the second funniest review in this data?

-- What is his/her name and how many reviews has he written?

-- Expected columns: name, review\_count

-- (15 pts) Q5: How many user have left tips to more business than reviews?

-- That is, for each user in this set, more businesses are getting tips from them than getting review from them.

-- Expected columns: user\_count

-- (20 pts) Q6: We define "boring people" as -

-- users have left reviews but...

-- none of the reviews were marked as useful, funny, or cool...

-- and he/she has left no tips ever

-- How many "boring people" do we see in the data?

-- (20 pts) Q7: What is the distribution of reviews per user?

-- Please generate the histogram of this.

-- What is the distribution look like? How do you describe it?

-- Is mean or median a better representation of the distribution?

-- Expected columns: review\_count, user\_count