Q1:
A:
(i) I agree with Ronald, because we don't have much information about how they sampled the data. The sample method would affect a lot. For example, if the variance of donation amount is large, the randomness of sampling would lead to much difference. Therefore, this doesn't prove anything if the sample space is not large enough or the sampling is not so random.

(ii) Using the data to run one-sided t-test with bin size 200 gives the p-value 0.999969 from which we can conclude that DEM's claim is reasonable. Based on the t-test result, DEM's claim is very likely to be right.

(iii)
(iv) The sampel size of each state varies a lot which makes them unequal from the perspective of statistics. However, in my solution, we treat each state equally, regradless of weight. This could lead to weakness of the conclusion.
(v)

Q2:
A:
(i)
(ii)

Q3: Probability and conditional probability.

(a) The Internet is a wonderful source of information about symptoms of rare diseases. Are you sneezing? It could be the West Nile virus! The West Nile virus (WNV) infected approximately 2,000 people in the United States last year1. Sheldon, your hypochondriac friend, is sneezing and heard about the West Nile virus on Twitter. He demands a test for the West Nile virus, why not? The test correctly identifies the presence of WNV in 95% of cases and only gives false positives in 1/10,000 cases. Unfortunately, the test indicates came back positive for West Nile virus and Sheldon is very concerned. Assume that in the population of the United States there are 300 million people susceptible to WNV.

   (i) What is the probability that Sheldon has WNV?

   (ii) The WNV virus is fatal in 5% of the cases. What is the probability that Sheldon will die this year? Assume a fatality rate of any cause (car accident, etc.) of 0.1%.

(b) Alice and Bob are playing a simple dice game. Each rolls one dice and the one with higher number wins. If the numbers are the same, they roll again. If Alice just won, what is the probability that she rolled a '4'?

A:
(a)

   (i)

$$P(true|positive) = 0.05956$$

   (ii)

$$P(die) = P(die\ because\ of\ WNV) + P(die\ because\ of\ other\ causes)$$
$$= 0.05956 * 0.05 + 0.001$$
$$= 0.003978$$

(b)

$$P('4'|won) = \frac{P(won|'4') * P('4')}{P(won)}$$
$$= \frac{1/2 * 1/6}{15/36}$$
$$= 1/5$$

Q4: (a) Plot the empirical complementary cumulative distribution (ECCDF) of comic characters appearances in comic books. The ECCDF $P[X > x]$ is defined as the fraction of characters with more than x comic book appearances. For instance, if superman appears in 1000 comic books and there are only 10 characters with more than 1000 comic book appearances out of 2000 characters, then $P[X > 1000] = 10/2000$.

IMPORTANT: Your plot should be in log-log scale.

(b) Let A be the adjacency matrix connecting characters to comic books, where $A_{i,j}$ has character i appearing on comic book j. Let A^T be the transpose of matrix A.

(i) What does $W = AA^T$ represent? Give the name of the entity with the largest degree in the graph that has adjacency matrix W?

(ii) What does $U = A^TA$ represent? Give the name of the entity with the largest degree in the graph that has adjacency matrix U?

(iii) Choose the correct option: If $U = A^TA$, then (1) $U^T = AA$, (2) $U^T = A^TA$, (3) $U^T = AA^T$, or (4) $U^T = A^TA^T$.

(iv) Let $P = D^{-1}W$, where $W = AA^T$ and D is a diagonal matrix where $D_{i,i} = \sum_j W_{i,j}$. Find the eigenvector x such that $x = Px$ and $\|x\|^2 = 1$, where $\|x\|^2 = \langle x, x \rangle$ is the inner product of x with itself.

A:
(a)

see appendix.

(b)

(i) W represents how many times character i and character j appear in the same book.

Captain American has the largest degree.

(ii) U represents how many characters book i and book j share.

COC 1 has the largest degree.

(iii) (2) is correct because matrix U is symmetric.

(iv)

$$x = Px \tag{1}$$

$$Dx = Wx \tag{2}$$

$$(Dx)_i = \sum_{j=1}^{n}(W_{ij}x_j) \tag{3}$$

$$D_{i,i}x_i = \sum_{j=1}^{n}(W_{ij}x_j) \tag{4}$$

$$(\sum_j W_{ij})x_i = \sum_{j=1}^{n}(W_{ij}x_j) \tag{5}$$

$$x_i = x_j \; for \; all \; i \; and \; j \tag{6}$$

$$x_i = \frac{1}{\sqrt{n}}, \; i = 1, 2, ..., n \tag{7}$$