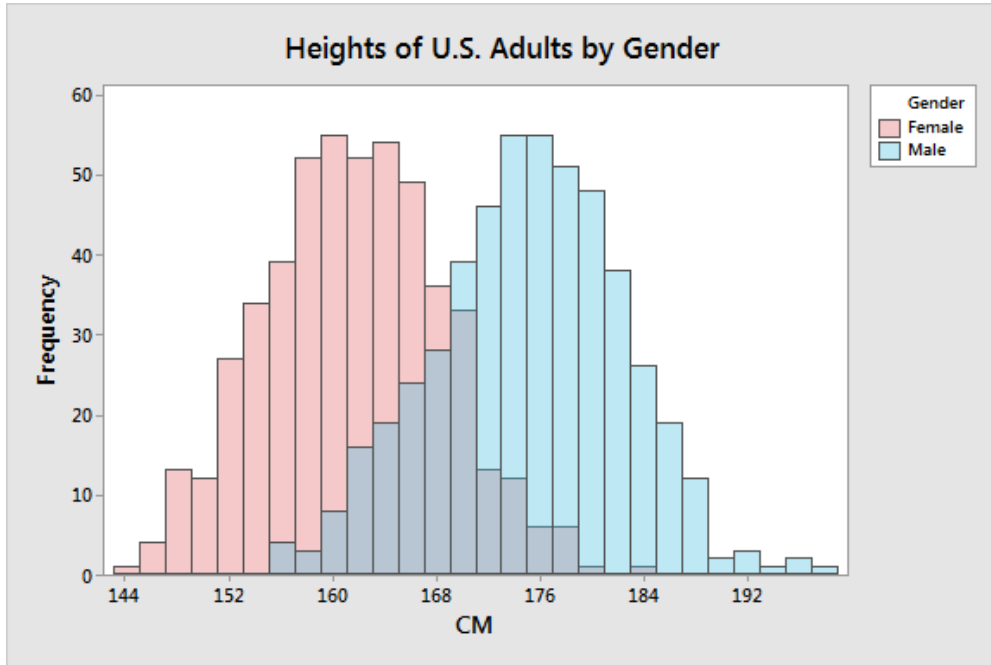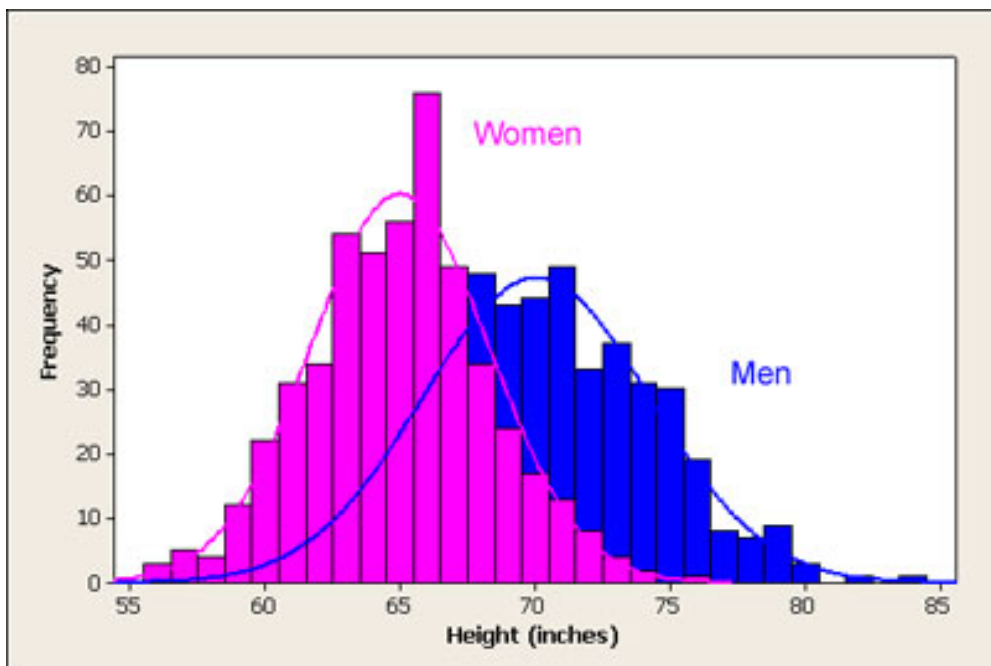# Illustrative Histograms and Distributions for Discussion
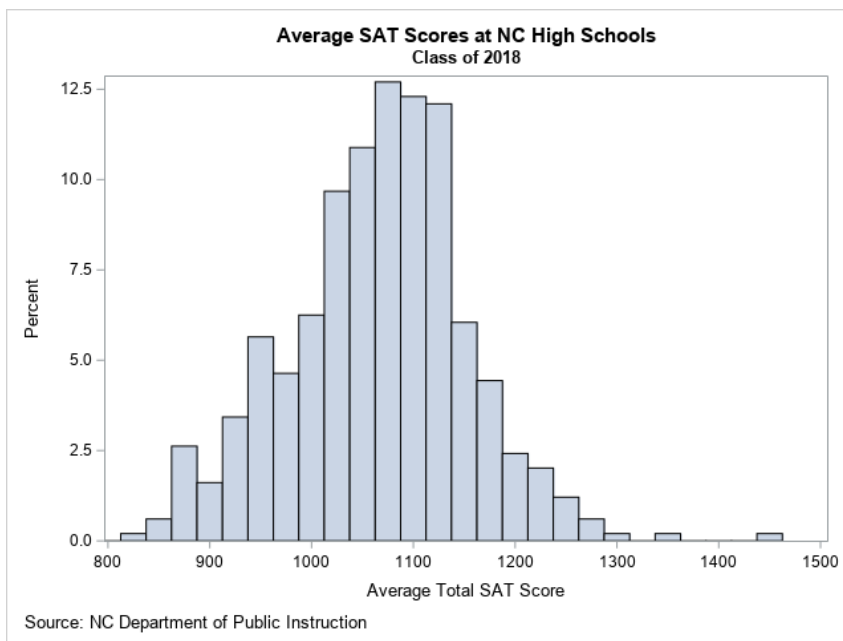
Illustrative Histograms.   Now we present and discuss a few illustrative histograms.

This is called a bi-modal histogram, having two peaks. Discussion how one average computed for everyone is misleading. We need two averages, one for each gender. Every time we come across a bi-modal distribution, we should think of it as displaying the results for two separate populations, and compute averages for each.

Sometimes it is useful to display a histogram together with the underlying model (distribution) that is generating the data. Here we can see that for heights of adults by gender. In each case, the distribution is a normal curve with parameters adjusted to be a best fit for the data.

Average SAT Scores at NC High Schools
Class of 2018

Source: NC Department of Public Instruction

This histogram shows the distribution of the **average total SAT score for <u>schools</u>** in North Carolina. These are not individual student scores, but average scores by school. That is why the right-hand tail and left-hand tail do not contain entries that you might expect if these were scores of individual students. Always be careful to understand exactly what is being shown.

From this histogram, you can determine several facts about the data:

1. For most NC schools, the average school-wide SAT score is about 1100.

2. About 73% of NC schools have an average SAT score between 1000 and 1200.

3. There are a few schools that have much higher scores than the others. Those schools are Early College At Guilford (Total=1442), Raleigh Charter High School (Total=1356), and East Chapel Hill High (Total=1290).

Can someone from the class sketch what the histogram might look like for individual student SAT scores from Early College At Guilford? Remember, this average is 1442!

https://blogs.sas.com/content/iml/2019/03/04/visualize-sat-scores-nc.html
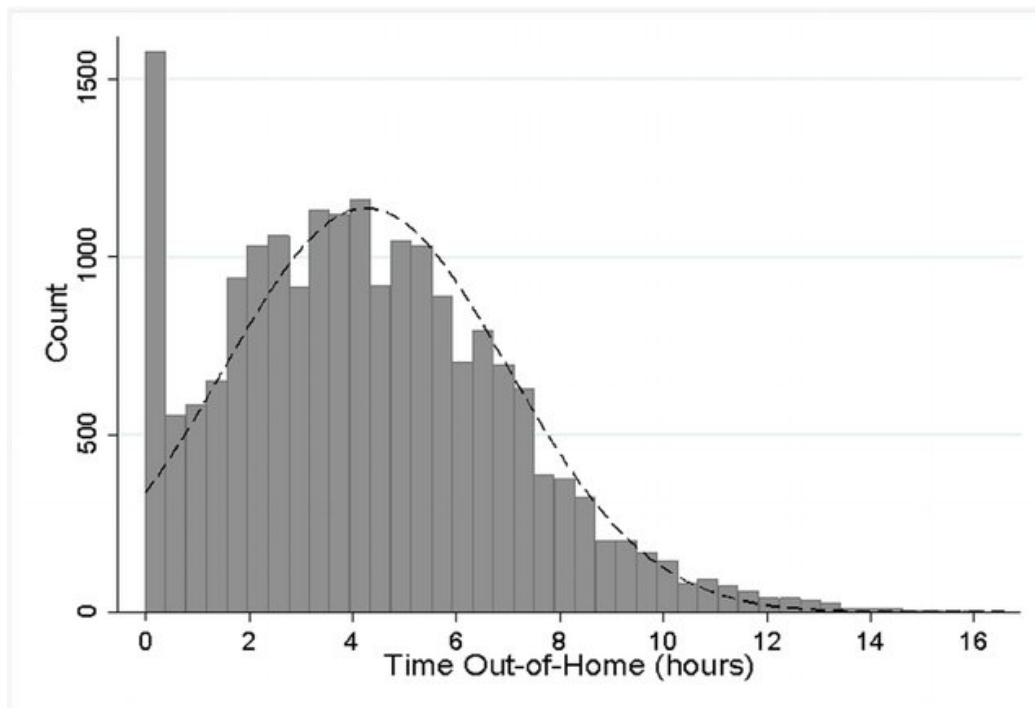
For older adults, histogram of the daily hours spent outside the home, showing the limit at zero. A normal distribution curve is plotted as a dashed line to show the data is approximately normally distributed except at and below zero. Here we have two separate populations displayed: Those who do get out of home and those who do not. Want to compute average for each, with one average equaling zero.

From *Time Out-of-Home and Cognitive, Physical, and Emotional Wellbeing of Older Adults: A Longitudinal Mixed Effects Model*

https://bit.ly/3hyAcXJ
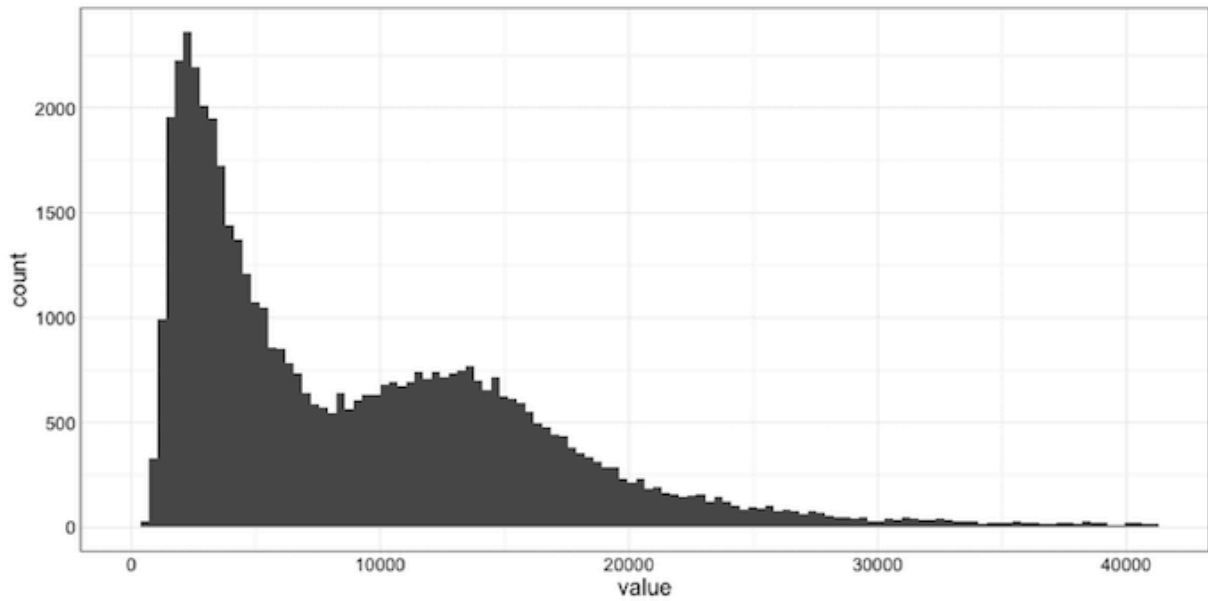
Here we are seeing the histogram of response times to various internet requests. Again, we see a multi-modal distribution, suggesting that at least two different populations are combined into one display. Separate averages are required, as the global average for this histogram is misleading – not giving useful information about either subpopulation.

https://blog.newrelic.com/engineering/expected-distributions-website-response-times/