1. What is a probability density function? What is a cumulative distribution function? Why are they important?

Using the Galton heigh data (<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/T0HSJ1> ) answer the following questions:

1. Create two histograms in a subplot (vertical or horizontal is fine) that show the distribution of height of children by gender.
   1. How do you choose your bin size?
   2. How is this data distributed?
2. Create a box and whisker plot for heights of children and heights of parents. What does this show you?
3. What is the probability that a person is taller than 78 inches?
   1. What do you notice about these records? Hint: look at the other columns.
4. Calculate the sample variance and standard deviation for all the numerical columns in the same output.
   1. Why sample variance and not population variance?
   2. What is your takeaway for each metric?
   3. What are the scales of measurement for each of these values?
5. What percentile is a child who is 75 inches in?
   1. 55 inches?
6. What minimum height are 80% of males taller than?
7. Explain The Monty Hall problem in your notebook in markdown in under 250 words.