1. The typical length of a hydrologic unit along Summit Cr is about 10m. The shortest is about 1m; the longest about 29m. The range of the values is thus around 28m (29 – 1 = 28).
2. Using the stripchart command, no information has been lost, though some of it is difficult to see as the boxes overlap one another. This issue is corrected using the stacked method. The pattern that sticks out in the stripchart is that a lot of the values are clustered between about 4m and 14—this was also evident in the first scatterplot, though in my opinion it can be seen faster and easier in the stripchart.
3. The advantage of the boxplot over the scatterplot is that the data is split into clear values, showing the median, IQR, minimum, and maximum. The scatterplot doesn’t automatically show these and the user would have to go through the data in order to decipher these. Using the boxplot, this work does not need to be done as the program displays these values for you.
4. The range that occurs most frequently in the histogram of PctObm is 0.5 – 0.6. The overall range of the values in the dataset is 0.2 – 1.0.
5. The distribution of PctObm is much more difficult to see in the stripchart than the histogram. From the stripchart, I would venture to guess based purely upon observation that the range occurring most frequently could either be between 0.4 – 0.5 or 0.5 – 0.6. Due to the overlapping of the boxes within the chart, it’s hard to visualize the actual count, which is given in the histogram.
6. After creating the histogram with breaks=20, the frequency changes from 15 to 12 and the range of bars appears to spread out, though there is not true change within the data. The number of bars increases as each point on the x-axis is split from maintaining a .1 range to a .05 range. With breaks=5, the histogram returns to maintaining bars that extend along the x-axis at a .1 range.
7. The density line has a similar shape to that of the histogram, but it’s less clear where exactly the data is distributed. The histogram reports this nicely with fitted bands in each range. The density line follows the trend of the data, meaning that this view seems to be more dependent upon the way that the plot is generated since the histogram simply takes the face value of the data points themselves. Because of this, the histogram view loses the least amount of information about the individual values.
8. The resulting plot contains the histogram limited to a density of 5 (y-axis) with each bar covering a 0.05 range of data. The lines() function adds in a trendline of the data over the top of the initial histogram, and the rug() function appears similar to the stripchart below the histogram showing the distribution of the data.