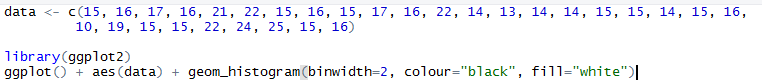
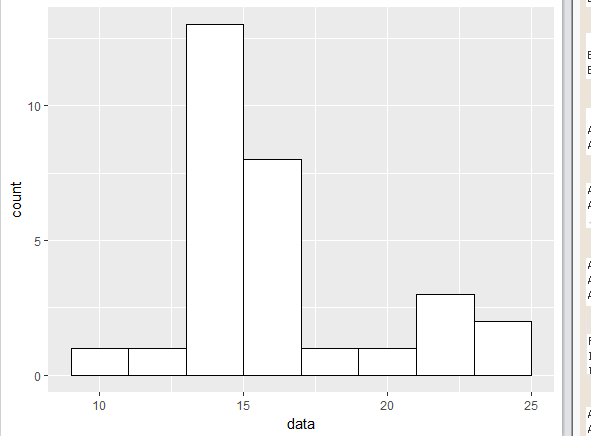
***Udacity Data Analyst Track***

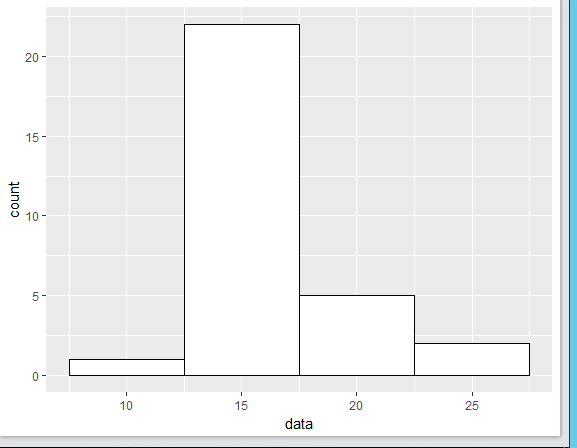
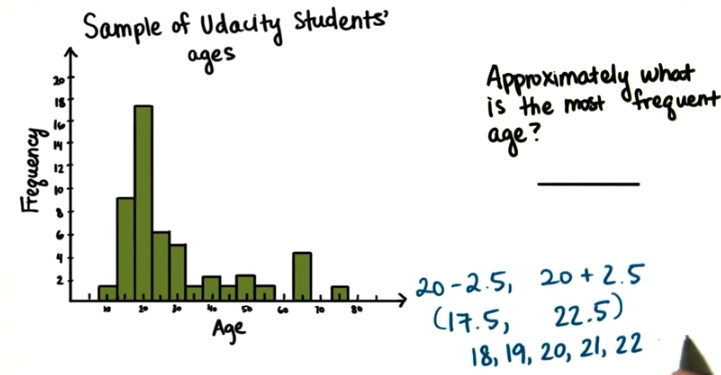
**I. Into to Descriptive Stats**

3. Central Tendancy

* **Frequency** = # of times a certain outcome occurs.
* **Proportion (RELATIVE frequency)** = the fraction of counts over a total sample (turned into a % by multiplying by 100)
* Are always between 0 and 1 (0% and 100%)
* **Histograms** = a graphical representation of the distribution of data where discrete intervals (**bins**) are decided upon to form widths for the boxes.
* Bin size is also a part of the frequency table 🡪 helps determine how many rows (bins of 5, ins of 10, 1 row per record, etc.)
* **Positive Skew** = outliers are present along the right-most end (tail) of the distribution
* **Negative Skew** = outliers are present along the left-most end (tail) of the distribution
* Kathleen counts the # of petals on all the flowers in her garden





* 13-15 petals seems most prominent in Kathleen’s garden. What happens
* if we change the bin size to 5 (too large bins = harder to see distribution 🡪 sometimes sacrifice detail for convenience)
* 
* The right skew in Kathleen’s flower petal distribution seem to indicate she has more shorter flowers, with few very tall flowers
* 
* If we have data from 15-105, if we have a bin size of 10, we end up with (105-15)/10 = 90/10 = 9 bins
* W/ a frequency table, we have exact counts (better to calculate n) so we can always create a histogram (better for analyzing distribution), but this is not vice versa