***Udacity Data Analyst Track***

**I. Into to Descriptive Stats**

2. Visualizing Data

* **Frequency** = # of times a certain outcome occurs.
* **Proportion** = the fraction of counts over a total sample (turned into a % by multiplying by 100)
* **Histograms** = a graphical representation of the distribution of data, discrete intervals (bins) are decided upon to form widths for our boxes. 8 Visualizing Data R Adjusting the bin size of a histogram will compact (or spread out) the distribution. Figure 2.1: histogram of data set with bin size 1 Figure 2.2: histogram of data set with bin size 2 Figure 2.3: histogram of data set with bin size 5 2.2.1 Skewed Distribution Definition 2.4 — Positive Skew. A positive skew is when outliers are present along the right most end of the distribution Definition 2.5 — Negative Skew. A negative skew is when outliers are present along the left most end of the distribution 2.2 Histograms 9 Figure 2.4: positive skew Figure 2.5: negative skew 10 Visualizing Data 2.3 Practice Problems Problem 2.1 Kathleen counts the number of petals on all the flowers in her garden, create a histogram and describe the distribution of flower petals on Kathleen’s flowers. Use a bin size of 2. 15 16 17 16 21 22 15 16 15 17 16 22 14 13 14 14 15 15 14 15 16 10 19 15 15 22 24 25 15 16 Table 2.1: Kathleens petal counts Problem 2.2 What number of petals seems most prominent in Kathleen’s garden? What happens if we change the bin size to 5? Problem 2.3 What does the skew in Kathleen’s flower petal distribution seem to indicate?