1. A concern of forest and conservation managers is the sale of forest lands that are then converted to non-forest uses (e.g. agriculture, residential, commercial, etc.). 775 “quarter sections” in a portion of northern Wisconsin were selected, and the number of acres of forest land converted to non-forest use during the period 1990-1999 was measured for each (a quarter section is a square region measuring 0.5 mile on all sides, equal to 160 acres). Below is the distribution table for forest loss:

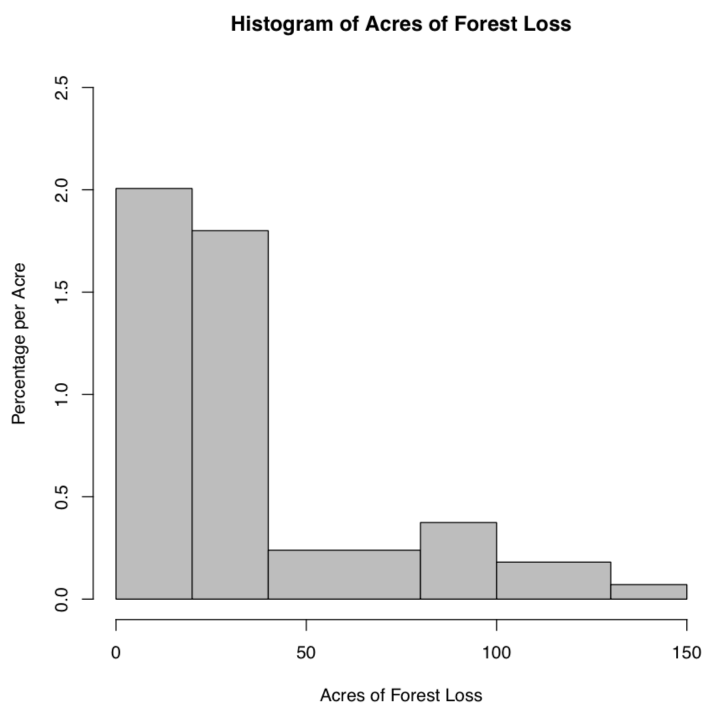
| Forest loss (acres) | 0-20 | 20-40 | 40-80 | 80-100 | 100-130 | 130-150 |
|---|---|---|---|---|---|---|
| # of quarter sections | 311 | 279 | 74 | 58 | 42 | 11 |

(a) Identify the subject, variable, and data type of the variable being studied.

*Each quarter section is a subject and for each quarter section, the number of acres of forest land converted to non-forest use is the variable, the data type of the variable is quantitative (continuous) since it is a numerical value that can be measured to whatever level of precision desired (tenth or hundredth or thousandths of an acre could be recorded).*

(b) Create the [density] histogram of forest loss. Don’t forget to label axes.

| Bin | Count | Percentage | Bin Length | Density= %/acre |
|---|---|---|---|---|
| [0,20) | 311 | 311/775=0.401 | 20 | 40.1/20= 2.0 |
| [20,40) | 279 | 279/775=0.36 | 20 | 36/20= 1.8 |
| [40,80) | 74 | 74/775=0.095 | 40 | 9.5/40=0.24 |
| [80,100) | 58 | 58/775=0.0748 | 20 | 07.48/20=0.374 |
| [100,130) | 42 | 42/775=0.0542 | 30 | 5.42/30=0.18 |
| [130,150) | 11 | 11/775=0.014 | 20 | 1.4/20=0.07 |




(c) Approximate the percentage of quarter sections that experienced forest loss of less than 80 acres.

*(311+279+74)/775=664/775=0.857 85.7% Here we will take the sections that have less then 80 acres, add their counts up and divide by the total count.*

(d) Estimate the percentage of quarter sections with forest loss between 20 and 50 acres.
*We will use all of 20-40 bin since this bin is completly contained in our target (20-50): 20-40: 36% . Next, we want to go from 40 to 50. But we don't have a bin this size, so we want to estimate. We look to see that the next bin is 40 to 80. We want the first 25 %. So we estimate that 25% of the 40-80 bin is in the 40-50 segment. 40-50: 0.095 $\times$ .25=2.4%, so total is 38.4 %*

(e) Give a possible range for the median number of acres converted to non-forest use.

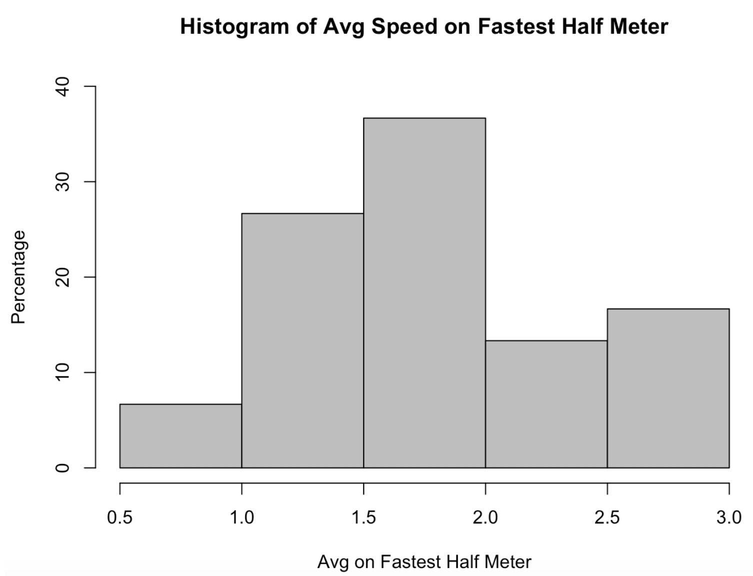
*Since there are 775 values, our median is the 388th value (387+1+387). Since 311+77=388, the median value occurs in the 20-40 bin.*

(f) Would you predict the mean number of acres converted to non-forest use to be larger than, smaller than, or have similar value to the median you found above? Why? Why can't you calculate the exact mean?
*I would predict the mean to be higher than the median since the histogram is skewed right. The mean will be pulled up towards the higher values. The exact mean cannot be calculated because we don't have the exact value for each subject.*

2. A zoologist collected wild lizards in the Southwestern United States. Thirty lizards from the genus Phrynosoma were placed on a treadmill and their speeds were measured. The recorded speed (meters/second) is each lizard's average speed on its fastest half meter.

0.50, 0.76, 1.04, 1.20, 1.24, 1.29, 1.28, 1.02, 1.36, 1.49, 1.55, 1.56, 1.57, 1.63, 1.57
1.94, 1.70, 1.72, 1.78, 1.78, 1.92, 2.11, 2.52, 2.10, 2.17, 2.54, 2.47, 2.57, 2.67, 2.66

A relative frequency histogram of the data with the following class intervals: [0.5, 1), [1, 1.5), [1.5, 2), [2, 2.5), [2.5,3) is given below. (notice the units of the y-axis)




(a) Based on the graph, what percent of the lizards averaged between 1 and 3 meters per second on its fastest half meter?
*Here we only need to subtract the percentage of the first interval, which is approximately 7, so 93 % between 1 and 3. Being clever and taking a second to think about the problem can very often save you alot of calcuations in these kind of scenarios.*

(b) Based on the graph, what percent of the lizards averaged above 2.25 meters per second on its fastest half meter?

*Simlair to before, 2.25-2.5: half of 13=6.5 % + 17 % for 2.5-3 = 23.5 %*

(c) Use the values given to calculate the mean of the lizards' average speeds on their fastest half meters.

*1.72 to calculate the mean, we sum up all the data points and then divided by 30*

(d) Use the values given to calculate the median of the lizards' average speeds on their fastest half meters.

*1.665, we order the data, then take the average of the 15th and 16th since 30 data points (1.63+1.70)/2=1.665. We take an average since there are even number of points.*

(e) Describe the relationship between the mean and median and how that is related to the shape of the graph.

*The values are close and fall into the same class. The graph is roughly symmetric so not surprised that they are similar.*

(f) While reviewing her notes, the zoologist noticed the data value she initially recorded as 2.67 should have been 2.76. Which of the measures of center and spread will she need to recalculate and which will she not? *Need to adjust mean but not median. Since this data point is not one of the two data points in the middle, we do not need to adjust median. Need to recalculate SD (reflective of all data points) and Range (since highest value), but not IQR since Q1 (1.308) and Q3 (2.107) do not change - if have not covered IQR by this point, just skip mentioning it for now.*

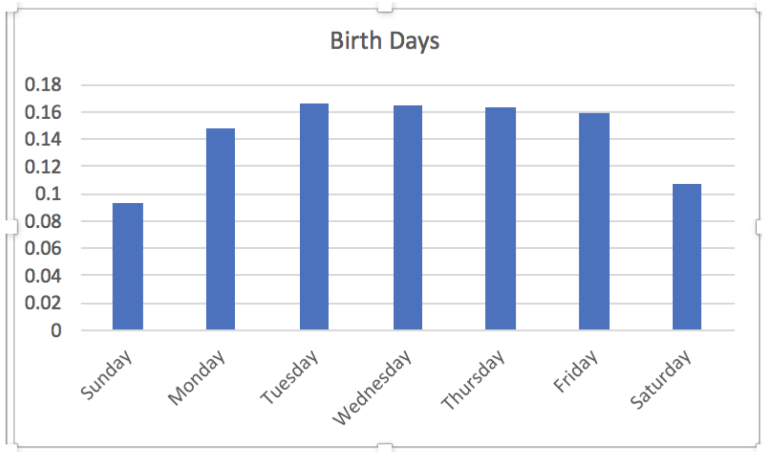
3. Births are not evenly distributed across the days of the week. Here are the average numbers of babies born on each day of the week at a hosptial in the United States in a recent year.

| Day | Births |
|---|---|
| Sunday | 7374 |
| Monday | 11704 |
| Tuesday | 13169 |
| Wednesday | 13038 |
| Thursday | 13013 |
| Friday | 12664 |
| Saturday | 8459 |

(a) What type of graph is useful to display the data? Create the graph. What patterns do the data show?

*Bar Graph since we have counts of categorical data (Mon, Tues, Wed, etc.); similar rates of birth days on week days with lowest rate on Sundays*

| Day | Births |
|---|---|
| Sunday | 7374/79421=0.093 |
| Monday | 11704/79421=0.147 |
| Tuesday | 13169/79421=0.166 |
| Wednesday | 13038/79421=0.164 |
| Thursday | 13013/79421=0.164 |
| Friday | 12664/79421=0.159 |
| Saturday | 8459/79421=0.107 |



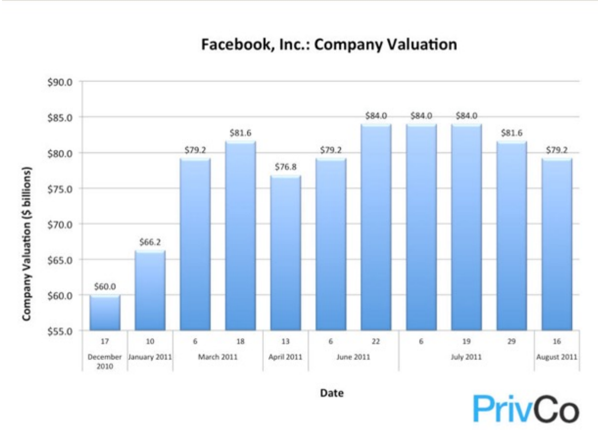

(b) Suggest some possible reasons why a lower percentage of births are on weekends.
*C-sections not typically scheduled for weekends.*

(c) Would it be appropriate to make a pie chart of these data? Explain.
*Yes, pie chart would be ok since these categories exhaust when births can happen and totals across these categories is the total number of births. By exhaust, we mean that by graphing all 7 days, we cover every possible outcome. With a pie chart, it may be difficult to compare values - determine which is smallest/largest since they are all very similar.*

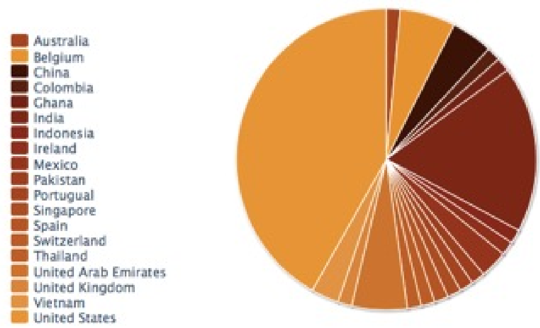
(d) Would it be appropriate to talk about the shape of these data? Explain.
*No. Since this is categorical data, shape doesn't mean anything; the order of the bars is arbitrary. I could have easily put Saturday and Sunday together and that would have changed the look of the distribution.*

4. Comment on the graphs below:

(a) Valuation of Facebook over time *1. Y-axis doesn't start at 0, makes "small" fluctuations look large. 2. Why did they choose the specific day points that they did to graph? Seems arbitrary.*




(b) Reported country of residence of participants at University of Chicago's summer scholars program given in a pie chart.*1. Many slices makes this unreadable. The colors and sizes of the slices are too similar to tell the difference between slices.*




(c) Stacked Bar Chart of Commute Times *1. This is a stacked bar-chart. Total height=100%. Larger area for cities with 3 letter abbreviations makes it harder to compare vertical percentages. Also hard to compare vertical distances of those cities that are not side-by-side. For example, it's hard to tell whether Chicago is more similar to DC or Seattle.*