1 Histograms

Key Points about Histograms

- The **area** of each bar in the histogram represents the proportion of the total population which is found in the specific range
- Area of a rectangle = $Width * Height$ in a histogram (and for all rectangles)
- Area is in proportion, width is in a certain specified unit (e.g. square inches), therefore, height (which equals $\frac{area}{width}$) has units proportion per whatever the unit is on the x-axis (e.g proportion per square inch).
- The sum of the areas should add up to 1, since they are all proportions
- Sometimes, people will use percentages instead of proportions for areas/height. In this case, everything on the y-axis would be multiplied by 100 and the areas should add up to 100 (instead of 1).

1. We have a group of mice, each of which has a specific weight. The distributions of weights is shown in the table below:

0lb-.5lb	.5lb-1.5 lb	1.5lb-2lb	2lb-5lb	6lb-12lb
10%	15%	20%	30%	25%

(a) Draw out the histogram in terms of proportion. What are the units of the y-axis?
We are asked to draw out the histograms in terms of proportions. So, we should convert all of the percentages in the table to proportions by dividing by 100. Our new table looks like this:

0lb-.5lb	.5lb-1.5 lb	1.5lb-2lb	2lb-5lb	6lb-12lb
.1	.15	.2	.3	.25

Now, we have widths, and we know that proportions should be the area of the bars. So, let's calculate the height for each of the bars:
Bar from 0-.5lb: Height $= \frac{Area}{Width} = \frac{.1}{.5} = .2$
Bar from .5-1.5lb: Height $= \frac{Area}{Width} = \frac{.15}{1} = .15$
Bar from 1.5-2lb: Height $= \frac{Area}{Width} = \frac{.2}{.5} = .4$
Bar from 2-5lb: Height $= \frac{Area}{Width} = \frac{.3}{3} = .1$
Bar from 6-12lb: Height $= \frac{Area}{Width} = \frac{.25}{6} = .04$
What are the units for height? The units for area are proportion of mice, the units for width are lb, so dividing the two, we get that the units for the y axis is proportion of mice per lb.

(b) Assume we lose the percentage of mice within the range of .5lb-1.5lb. How can we recover it?
Let's work with the proportions table from earlier. We know that the proportions should add up to 1, so we can take all of the proportions we know, sum them up, and figure out how far away we are from 1 .In this case, we have that proportion of mice within the range of .5-1.5 $= 1-(.1+.2+.3+.25) = .15$. The question, however, asks for percentage. To get percentage from proportion, we multiply by 100, so our percentage is $.15*100 = 15\%$
Note: We could have worked with the original percentages table and subtracted from 100 rather than 1; this is just an example to show that they are essentially equivalent, despite this factor of 100.

(c) Re-draw the histogram in terms of percentages instead of proportions. What are the units of the y-axis now?
What's different? Now, our areas should not add up to 1, but rather, they should add up to 100. So, our area of each bar is now our proportion multiplied by 100. Luckily, these are just the numbers that were given to us in our original table!
Bar from 0-.5lb: Height $= \frac{Area}{Width} = \frac{10}{.5} = 20$
Bar from .5-1.5lb: Height $= \frac{Area}{Width} = \frac{15}{1} = 15$
Bar from 1.5-2lb: Height $= \frac{Area}{Width} = \frac{20}{.5} = 40$
Bar from 2-5lb: Height $= \frac{Area}{Width} = \frac{30}{3} = 10$
Bar from 6-12lb: Height $= \frac{Area}{Width} = \frac{25}{6} = 4$
The units of height are now area (which is in percentage of mice) over width (still in lb), which are percentage of mice per lb.

2. Different students at Cal pay different amounts of money for their textbooks. Assume we create a valid histogram displaying the distribution of money payed, relating it with how much of the student population pays within that range. Below is the range of money, and what the height of the bars of histograms is for each range.

\$0-\$50	\$50-\$75	\$75-\$150	\$150-\$300
.0075	.002	x	y

(a) Assume the units on the y-axis is in proportion of students per dollar. Moreover, assume that there are 20,000 students. Can you fill in the values for x and y? Why or why not.
No; we can calculate the areas of the first two bars, we know that the addition of all bars should add up to 1, but we are missing two bars, so we can not complete this calculation.

(b) With the assumptions above, also assume that 8,000 of them spend \$150-\$300. Which values can you now fill in? Calculate everything you can fill in.
Yes!
Let's note what this assumption tells us: $\frac{8000}{20000} = .4$ is the proportion of people in the 150-300 bin. We have proportion of the bar (which equals the area as shown by the y-axis), we have the width of the bar ($300 - 150 = 150$), so we can calculate the height of the bar! Height y $= \frac{Area}{Width} = \frac{.4}{150} = .0027$. This is our value of y.
How does this tell us the value of x? We know that all of the proportions add up to 1, and we can calculate the proportions of the three bars we do have given their widths and heights. From there, we will have the area of the 75-150 bar, and we can calculate the width as $150 - 75 = 75$, so we can go ahead and get the height.
Our missing proportion is $1 - (.4 + (.0075)(50) + (25)(.002)) = .175$
We have the proportion of the 75-150 bar, we have the width being 75, so dividing the proportion by width, $\frac{.175}{75} = .0023$, gets us x, our missing height!
Biggest advice for these problems is to take it one step at a time. You will get less confused this way.

(c) Repeat the above question, but now assuming the y-axis in in percentage of students per dollar. Which one makes more intuitive sense?
This new assumption tells us that we are working with percent rather than proportions. This means that things should add up to 100, not 1. Moreover, the area of our bars should represent percentages now.
.4 is the proportion of people in the 150-300 bin as shown above, so that means that the percentage of students is 40%.
Height y $= \frac{Area}{Width} = \frac{40}{150} = 27$. This is our value of y.

How does this tell us the value of x? We know that all of the percentages add up to 100, and we can calculate the percentage of the three bars we do have given their widths and heights. From there, we will have the area of the 75-150 bar, and we can calculate the width as $150-75=75$, so we can go ahead and get the height.
Our missing percentage is $100-(40+(.0075)(50)+(25)(.002))=59.75$
We have the percentage of the 75-150 bar, we have the width being 75, so dividing the percentage by width, $\frac{59.75}{75}=794$, gets us x, our missing height!
Which one makes more intuitive sense? Both methods say that there is 40 percent of students in the 150-300 bar, but this second method, assuming percentages, says that 59.75 percent of students are in the 75-150 bar. That means that the leftover percentage that is not in these two, .25 percent of students, are in the other two bars. This doesn't make much sense.
On the other hand, when we assumed proportions, roughly .175, or 17.5 percent of students were in this 75-150 bar, leaving a lot more room for students to be in the first two bars. In this way, it makes a bit more sense for the given heights to correspond to a proportion histogram rather than a percentage one.