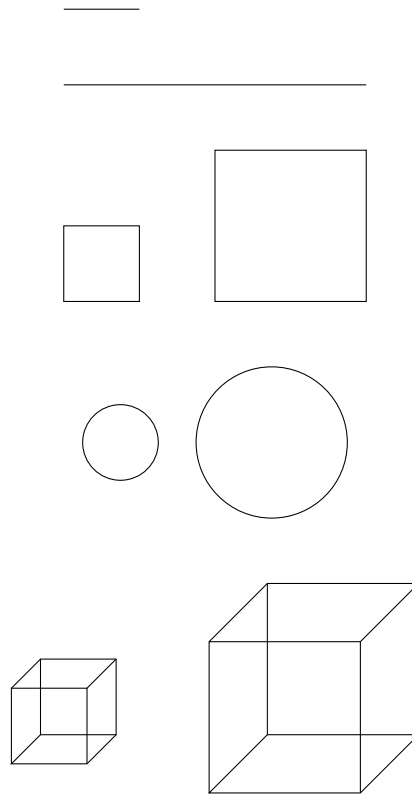# Data Visualization

Question 1: How many times is the bigger one bigger than the smaller one in terms of length, area, and volume?

Discuss how your estimates compare to the actual ratios. Did you notice that it became harder to estimate accurately for areas and volumes?

Question 2: Histograms

1. Draw a histogram representing the number of hours people spend watching Game of Thrones per week. Use the following data:

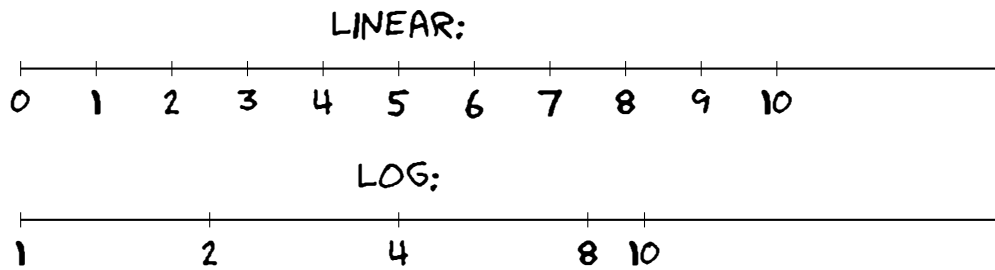| HOURS PER WEEK	NUMBER OF PEOPLE
$(0, 2]$	15
$(2, 4]$	25
$(4, 6]$	35
$(6, 8]$	20
$(8, 10]$	5

Use equal bin sizes and label your axes appropriately.

2. Now, draw another histogram using the same data but with varying bin sizes:

HOURS PER WEEK	NUMBER OF PEOPLE
$(0, 2]$	15
$(2, 4]$	40
$(4, 8]$	45
$(8, 16]$	30
$(16, 32]$	20

How does this change the appearance of the distribution?

Question 3: Consider the following linear and logarithmic scales: Mark the position of 5 and 25 on the logarithmic scale.

LINEAR:

```
├──┼──┼──┼──┼──┼──┼──┼──┼──┼──┼─────────
0  1  2  3  4  5  6  7  8  9  10
```

LOG:

```
├─────────┼──────────┼──────────┼──┼────
1         2          4          8  10
```

Question 4: Draw a histogram using a logarithmic scale for the x-axis (hours watched) using the data from Question 2.2.

Question 5:

The Cumulative Distribution Function (CDF) of a random variable X is defined as:
$$F(x) = P(X \leq x)$$

It represents the fraction of data points that are less than or equal to x. The Complementary Cumulative Distribution Function (CCDF) is defined as:
$$\bar{F}(x) = P(X > x) = 1 - F(x)$$

It represents the the fraction of data points that are greater than x.

1. Draw the CDF and CCDF of the following data in linear scale.

{0, 1, 2, 5, 3, 7, 1, 9, 4, 6, 2, 8, 5, 10, 5, 7, 4, 8, 8, 1}

2. Draw the CDF and CCDF of the following data in logarithmic scale (both x and y axes).

{1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 1, 2, 4, 8, 16, 32, 256, 1, 2, 4, 32, 64, 1}