EX 1.2.1: Given the following sample of college majors in a Linear Algebra course:

HIST, IE, CHE, ME, CS, IE, CHE, ME, MATH, MATH, IE, ME, ECO, ME, CHE, CS, MATH, ME, MATH, CS

a) Construct the frequency table for the data.

MAJOR	FREQUENCY	RELATIVE FREQUENCY (= DENSITY)
CHE	3	$3/20 = 0.15$
CS	3	$3/20 = 0.15$
ECO	1	$1/20 = 0.05$
HIST	1	$1/20 = 0.05$
IE	3	$3/20 = 0.15$
MATH	4	$4/20 = 0.20$
ME	5	$5/20 = 0.25$
TOTAL:	**20**	**$20/20 = 1.00$**

SANITY CHECKS: Total Frequency should equal # data points, Total Rel. Freq. should be very close to one

Here, the categories are listed in **alphabetical order**, which is preferred but **not required**.

Remember, for **categorical** data, **density** <u>**equals**</u> **relative frequency**.

b) Construct a histogram for the data with **density** as the vertical axis.

Histogram of college majors in a Linear Algebra class
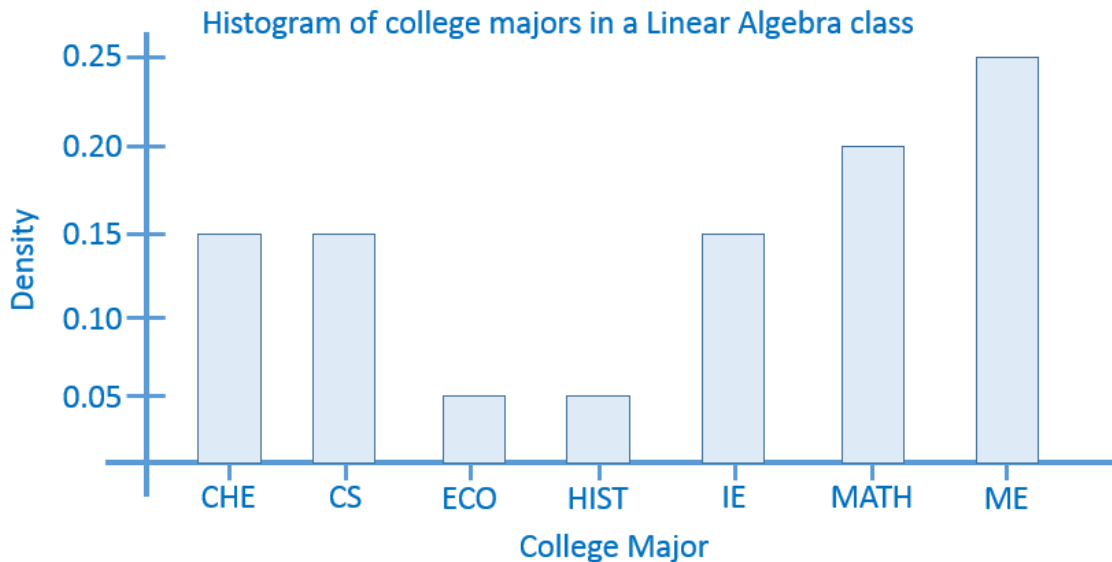
c) What **proportion** of the students are MATH majors?

Look at the **RELATIVE FREQUENCY** column: RelFreq[MATH] = $4/20$ OR 0.20

d) What **percent** of the students are Chemical Engineering (CHE) majors?

Look at the **RELATIVE FREQUENCY** column: RelFreq[MATH] $= 3/20 = 0.15 =$ 15%

e) Which major has the most students?

Use histogram to find the category with the **highest (relative) frequency**: Mechanical Engineering (ME)

EX 1.2.2: Given the following dataset of # strikes in completed bowling games: 2, 7, 3, 1, 2, 3, 6, 0, 12, 4, 3, 2, 3, 3, 4, 5

Moreover, suppose the **bin widths** (**class widths**) are **all equal** with width **one**, leading to the **bins** (**classes**):

0-<1, 1-<2, 2-<3, 3-<4, 4-<5, 5-<6, 6-<7, 7-<8, 8-<9, 9-<10, 10-<11, 11-<12, 12-<13

— OR IF YOU PREFER INTERVAL NOTATION —

$[0, 1)$, $[1, 2)$, $[2, 3)$, $[3, 4)$, $[4, 5)$, $[5, 6)$, $[6, 7)$, $[7, 8)$, $[8, 9)$, $[9, 10)$, $[10, 11)$, $[11, 12)$, $[12, 13)$

a) Construct the frequency table for the data.

BIN (# STRIKES/GAME)	FREQ.	REL. FREQ.	BIN WIDTH	DENSITY
$[0, 1)$	1	1/16	$1 - 0 = 1$	$1/16 \div 1 = 1/16 = 0.0625$
$[1, 2)$	1	1/16	$2 - 1 = 1$	$1/16 \div 1 = 1/16 = 0.0625$
$[2, 3)$	3	3/16	$3 - 2 = 1$	$3/16 \div 1 = 3/16 = 0.1875$
$[3, 4)$	5	5/16	$4 - 3 = 1$	$5/16 \div 1 = 5/16 = 0.3125$
$[4, 5)$	2	2/16	$5 - 4 = 1$	$2/16 \div 1 = 2/16 = 0.1250$
$[5, 6)$	1	1/16	$6 - 5 = 1$	$1/16 \div 1 = 1/16 = 0.0625$
$[6, 7)$	1	1/16	$7 - 6 = 1$	$1/16 \div 1 = 1/16 = 0.0625$
$[7, 8)$	1	1/16	$8 - 7 = 1$	$1/16 \div 1 = 1/16 = 0.0625$
$[8, 9)$	0	0/16	$9 - 8 = 1$	$0/16 \div 1 = 0/16 = \quad 0$
$[9, 10)$	0	0/16	$10 - 9 = 1$	$0/16 \div 1 = 0/16 = \quad 0$
$[10, 11)$	0	0/16	$11 - 10 = 1$	$0/16 \div 1 = 0/16 = \quad 0$
$[11, 12)$	0	0/16	$12 - 11 = 1$	$0/16 \div 1 = 0/16 = \quad 0$
$[12, 13)$	1	1/16	$13 - 12 = 1$	$1/16 \div 1 = 1/16 = 0.0625$
TOTAL:	**16**	**16/16 = 1.00**	**13**	**(IRRELEVANT)**

SANITY CHECKS: Total Frequency should equal # data points, Total Rel. Freq. should be exactly equal to one

SANITY CHECK: (**Lower bound** of **Bottom Bin**) + (**Total Bin Width**) = (**Upper Bound** of **Top Bin**)

Here, the **bottom bin** is $[0, 1)$ with a **lower bound** of **0** & the **top bin** is $[12, 13)$ with an **upper bound** of **13**.

b) Construct a histogram for the data with **density** as the vertical axis.

Histogram of bowling strikes
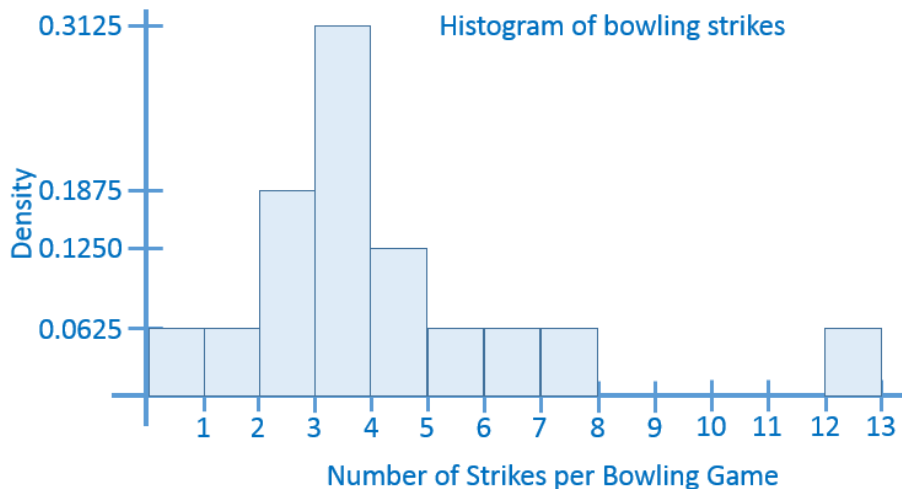
c) Describe the modality & skewness of the data.

Looking at the histogram, the data is ⌐unimodal⌐ and ⌐positively skewed⌐

d) Looking at the histogram, does there appear to be any outliers?   If so, identify them.

Yes, the **12 strikes** data point is an outlier as it's far away from the rest of the data.

EX 1.2.3: Given the following dataset of completed bowling games: 2, 7, 3, 1, 2, 3, 6, 0, 12, 4, 3, 2, 3, 3, 4, 5

Here, the characteristic observed for each bowling game is the **number of strikes**.

Moreover, suppose the **bin widths (class widths)** are **unequal** in way which leads to the following **bins (classes)**:

$$0\text{-}<1, \quad 1\text{-}<3, \quad 3\text{-}<6, \quad 6\text{-}<13$$

— OR IF YOU PREFER INTERVAL NOTATION —

$$[0,1), \quad [1,3), \quad [3,6), \quad [6,13)$$

a) Construct the frequency table for the data.

BIN (# STRIKES/GAME)	FREQ.	REL. FREQ.	BIN WIDTH	DENSITY
$[0,1)$	1	$1/16$	$1-0=1$	$1/16 \div 1 = 1/16 = 0.0625$
$[1,3)$	4	$4/16$	$3-1=2$	$4/16 \div 2 = 2/16 = 0.1250$
$[3,6)$	8	$8/16$	$6-3=3$	$8/16 \div 3 = 1/6 \approx 0.167$
$[6,13)$	3	$3/16$	$13-6=7$	$3/16 \div 7 = 3/112 \approx 0.027$
TOTAL:	**16**	$16/16 = 1.00$	**13**	**(IRRELEVANT)**

SANITY CHECKS: Total Frequency should equal # data points,   Total Rel. Freq. should be exactly equal to one

SANITY CHECK: (**Lower bound** of **Bottom Bin**) + (**Total Bin Width**) = (**Upper Bound** of **Top Bin**)

Here, the **bottom bin** is $[0,1)$ with a **lower bound** of **0** & the **top bin** is $[6,13)$ with an **upper bound** of **13**.

b) Construct a histogram for the data with **density** as the vertical axis.
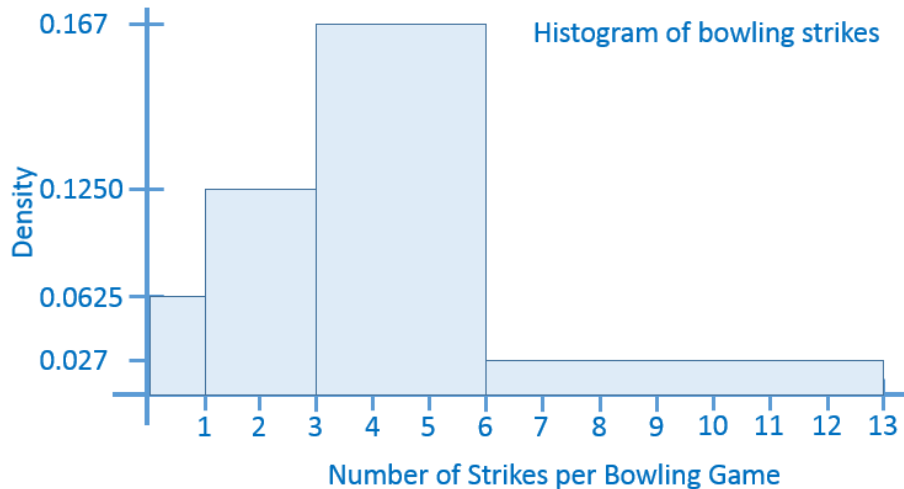
c) What proportion of the bowling games have at most five strikes?

Identify "uber-bin" corresponding to "at most five strikes": $[0,6)$

Decompose "uber-bin" as a **union** of the given bins: $[0,6) = [0,1) \cup [1,3) \cup [3,6)$

Add the corresponding relative frequencies: $1/16 + 4/16 + 8/16 = \boxed{13/16}$ OR $\boxed{0.8125}$

d) What percent of the bowling games have at least six strikes?

Identify "uber-bin" corresponding to "at least six strikes": $[6,13)$

In this case, the "uber-bin" is also one of the given bins.

Look up the bin's corresponding relative frequency: $3/16 = 0.1875 = \boxed{18.75\%}$