

Name \_\_\_\_\_ Period \_\_\_\_\_

### Skill 1.1 Exercise 1

Let's imagine we are working for the city government of the fictional city of Melody Metropolis. The mayor of Melody Metropolis wants to know more about the musicians who currently live in the city.

Open the dataset below and make a copy.

<https://docs.google.com/spreadsheets/d/1m9XbRZ40Deds2HUm4S04hFPjUywZpRfZMII2TbXIFic/edit?usp=sharing>

How would you describe this dataset? In your group, discuss the following,

- What does a typical musician's income look like?
- Is there a wide range of musician ages?
- What proportion of the musicians in the dataset play guitar?

How accurate are your answers to the questions above?

### Skill 1.2 Exercise 1

The variable instrument from our dataset tells us the type of instrument each musician plays. The table below gives the frequency, proportion, and percentage of musicians that play each type of instrument.

Fill in the missing entries to complete the table. Round all proportions to two decimal places and all percentages to the nearest whole number with the % symbol included after.

Instrument Categories			
INSTRUMENT	FREQUENCY	PROPORTION	PERCENTAGE
voice	262	0.27	
guitar	278		29%
piano	178		19%
drums	135	0.14	
saxophone	48		5%
violin	57	0.06	
<b>TOTAL</b>	<b>958</b>	<b>1.00</b>	<b>100%</b>

What is the ratio of voice to violin?

What is the ratio of guitar to drums?

Name \_\_\_\_\_ Period \_\_\_\_\_

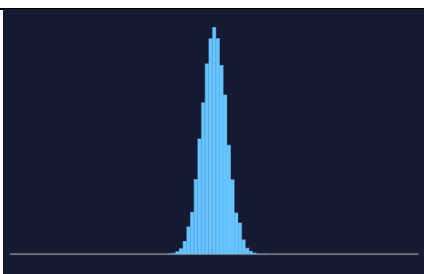
**Skill 1.3 Exercise 1**

Refer to the musician dataset. Which variables are categorical? Which are numerical?

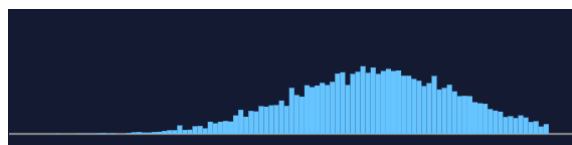
**Skill 1.4 Exercise 1**

Refer to the datasets below to answer the following,

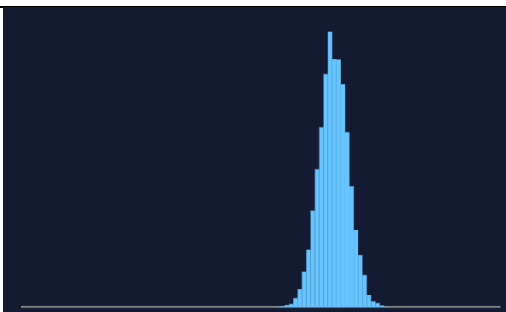
A.



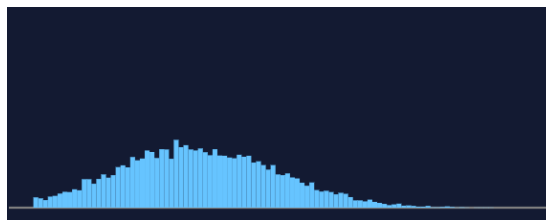
B.



C.



D.

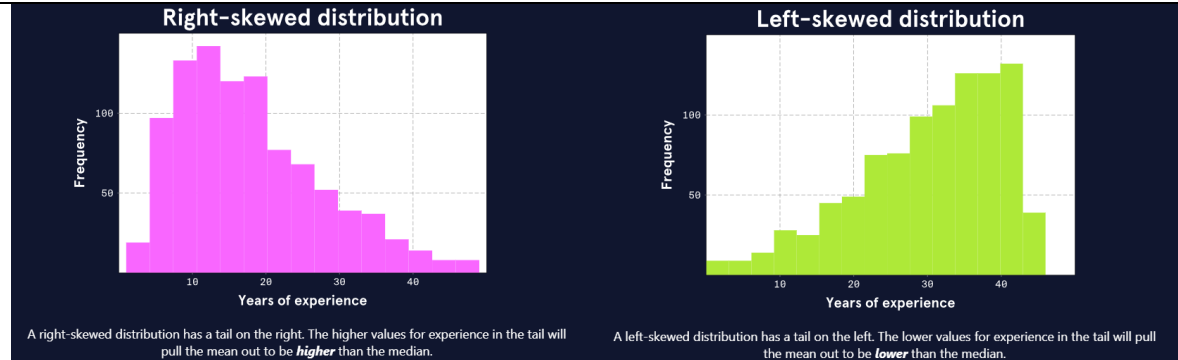


For which dataset(s) is/are the value of the mean the highest?

For which dataset(s) is/are the value of the standard deviation the greatest?

Name \_\_\_\_\_ Period \_\_\_\_\_

### Skill 1.5 Exercise 1



Another numeric variable from the musician dataset is years of experience working in the field of music. The plot on the left shows what the experience distribution might look like if it is right-skewed like the income distribution is. The plot on the right shows what the experience distribution might look like if it is left-skewed.

Refer to the musician dataset. Which distribution most likely is true for the musicians in Melody Metropolis?

### Skill 1.6 Exercise 1

Return to the musician dataset and navigate to the *Experience* tab. Do the following,

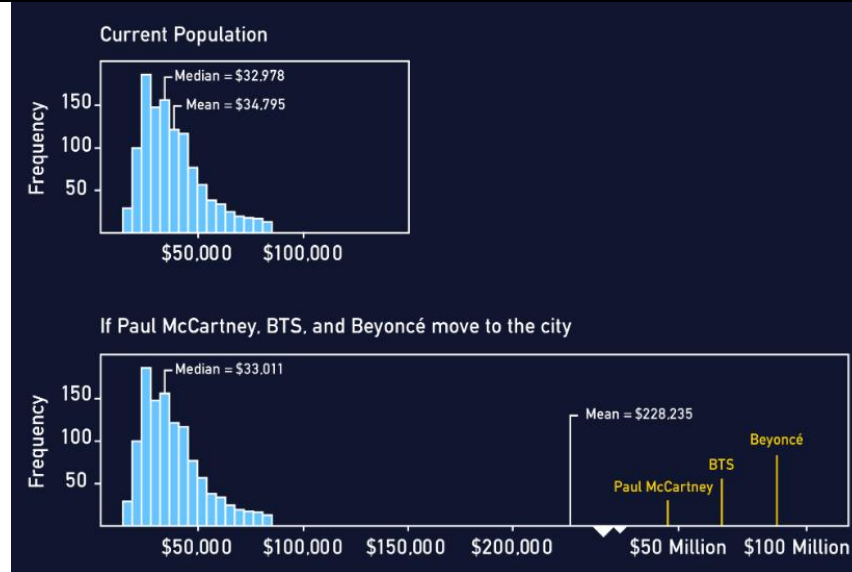
- Highlight all the data
- From the data menu, select “sort range” → “A to Z”
- From the data menu, select “Column stats” and record the following,

<b>Average</b>	
<b>Median (Q2)</b>	
<b>Median of all the values less than Q2 (Q1)</b>	
<b>Median of all the values greater than Q2 (Q3)</b>	
<b>Difference between Q1 and Q3 (IQR)</b>	

Are the years of experience right skewed or left skewed?

Name \_\_\_\_\_ Period \_\_\_\_\_

### Skill 1.7 Exercise 1



The second plot above shows that the median appears almost unaffected by the addition of these three gigantic incomes: the median moves from \$32,978 to \$33,011. However, the mean makes a drastic change from \$34,795 to \$228,235.

Which alternative statistics could your colleague use in this case?

### Skill 1.8 Exercise 1

Return to the musician dataset and navigate to the Experience and Instrument tab.

What is the average experience for each type of instrument?

Instrument	Average
Drums	
Guitar	
Piano	
Saxophone	
Violin	
Voice	

Name \_\_\_\_\_ Period \_\_\_\_\_

<b>Skill 1.9 Exercise 1</b>
Return to the musician dataset and navigate to the Age and Income tab. Do the following, <ul style="list-style-type: none"><li>• Highlight all the data</li><li>• From the Insert menu, select Chart</li></ul>
Is the correlation between these variables positive or negative? Is the correlation strong or weak?