1. Visualization Fundamentals

Visualization is an incredibly important tool in a data scientist's toolkit, enabling you to better understand the data you're working with and to share insights with others. However, not all visualizations are created equal — what **visualization type** works best for your data depends heavily on the **type(s) of data** you're working with and what you're trying to show with your visualization.

a. Describe what is meant by 'encoding' in the context of visualization.

Encoding describes how variables are represented in visual characteristics of a graph. A visualization 'encodes' one or more variables into one or more visual aspects of the plot, such as the x-coordinate, y-coordinate, length/height, area, shape, or color.

b. For each of the following variables, determine its variable type.

i. Phone number

Categorical Nominal

ii. Occupation (e.g. accountant, construction worker, etc.)

Categorical Nominal

iii. Day of the week

Categorical Ordinal

iv. Income

Numerical Continuous

v. Number of books owned

Numerical Discrete

c. Match the variable type(s) to the most appropriate visualization type.

1 Numerical Discrete Variable	
1 Numerical Continuous Variable	Bar Chart
1 Categorical Variable (Ordinal/Nominal)	Histogram
1 Categorical Variable, 1 Numerical Variable	Line Plot
2 Numerical Variables	Scatter Plot
2 Numerical Variables (one of which is time)	

2. Charts, Graphs and Plots Galore

`thai_restaurants` table:

Restaurant	Dish	Price ($)	Spiciness	Avg. Rating
Racha Cafe	Pad See Ew	10.95	4	4.55
Racha Cafe	Pad Thai	10.95	2	3.79
Imm Thai	Tom Yum Soup	7	3	4.09
Imm Thai	Pad Thai	14.5	1	4.12
Imm Thai	Spicy Fried Rice	13	5	4.81

(... 15 rows omitted)

Using the table `thai_restaurants` above, write code to create the following visualizations.

a. A histogram showing the distribution of prices across all dishes in the `thai_restaurants` table.

```
thai_restaurants.hist('Price ($)', density = False)
```

b. A histogram showing the price distribution of dishes, grouped by restaurant.

```
thai_restaurants.hist('Price ($)', group = 'Restaurant', density = False)
```

c. A bar chart showing the spiciness level for each dish across all restaurants.

```
thai_restaurants.barh('Dish', 'Spiciness')
```

d. A scatter plot showing the relationship between price and average rating, with different colors for each unique dish.

```
thai_restaurants.scatter('Price ($)', 'Avg.  Rating', group = 'Dish')
```

3. Interpreting Histograms

Histograms allow us to visualize the distribution of a single numerical variable by grouping numerical values into **bins** and encoding the **frequency** (count) of each bin as its height. While it is perfectly acceptable to combine histogram bins, you cannot split a bin as you don't know the distribution of values within a single bin.

Using the histogram above, answer the following questions.

a. What is the most common range of prices for dishes?

[10, 12)

b. Approximately how many dishes cost $14 or more?

13

c. True or False: Most dishes cost at least $10.

True

d. True or False: More dishes cost between $5 and $6 than between $4 and $5.

Cannot determine using the histogram — you can't split bins.