
Week 2 Tutoring

— Table manipulation and
visualization & histograms! —

Table Methods

Method	Function
<code>tbl.num_rows</code> <code>tbl.num_cols</code>	Returns number of rows or columns in the table
<code>tbl.column(name)</code>	Returns values of that column as an array
<code>tbl.select(col1, col2, ...)</code>	Returns table of only those columns *col1, col2 can be either column indices or column names (in the form of strings)
<code>tbl.drop(col1, col2, ...)</code>	Returns table without those columns
<code>tbl.sort(col_name)</code> <code>tbl.sort(col_name, descending=True)</code>	Returns table sorted based on specified column
<code>tbl.show(n)</code>	Returns first n rows of table
<code>tbl.take(row_indices)</code>	Returns table with only specified row index or indices (input as an array)

Table Methods (cont.)

Method	Function
<code>tbl.where(col_name, condition)</code>	Returns new table with all rows satisfying condition

*Note that the `tbl.where` method takes in a column name and a condition, and spits back all rows of the table that satisfy that condition. The condition can be a value (e.g. a string or an integer), in which case the method checks for equality. The condition can also be a predicate. See next slide for a list of predicates.

Table Predicates

Predicate	Description
<code>are.equal_to(Z)</code>	Equal to Z
<code>are.above(x)</code>	Greater than x
<code>are.above_or_equal_to(x)</code>	Greater than or equal to x
<code>are.below(x)</code>	Less than x
<code>are.below_or_equal_to(x)</code>	Less than or equal to x
<code>are.between(x, y)</code>	Greater than or equal to x, and less than y
<code>are.strictly_between(x, y)</code>	Greater than x and less than y
<code>are.between_or_equal_to(x, y)</code>	Greater than or equal to x, and less than or equal to y
<code>are.containing(S)</code>	Contains the string S

Table Visualization

- Relation between two numerical variables:
 - Scatter plots
 - Line graphs (often used to study change over time)
 - Both axes are numerical
- Distribution of one variable
 - Categorical: bar charts
 - Numerical values doesn't necessarily mean numerical variable
 - Remember census sex codes (0, 1, 2 for genders)
 - Numerical: histograms

Table Visualization Concept Check

1. You have collected data on the weights, heights, year of birth, and breeds of all dogs in your neighborhood. What kind(s) of data did you collect?
2. You want to see the relationship between dog weights and heights. What do you use to display your data?
3. What about for the distribution of years of birth?

Histograms

- Displays the distribution of a numerical variable
 - E.g. variable = ages
- Start with the axes!
- X-axis:
 - Bins are groupings of the numerical variable into continuous intervals
 - E.g. ages could be split into three bins: 0-45, 45-70, 70-120
 - Note that bin widths may not be equal
 - Bins are left inclusive, right exclusive - $[a, b)$

Histograms (cont.)

- Y-axis:
 - Not proportion of items in bin but **proportion of items relative to width of bin (density)**
 - From textbook: **area of bar = height of bar x width of bar**
 - So:
 - Height of bar = area of bar / width of bar
- Think of each bar's area as representing the % of items in the bin
- Why not just plot counts?
 - Unequal bin widths

Histograms - Conceptual Check

1. How do we find what percent of items fall within a bin?
2. What does the y-axis of a histogram measure?
3. Why don't we use counts for the y-axis for a histogram?
4. What does the area of a histogram bar represent?

Histograms - Numerical Check

Bin (height in cm)	0-70	70-100	100-190	190-200	200-250
Height (proportion per cm)	0.00343	0.007	0.0047	?	0.0002

Given the above table of bin widths and heights for a particular histogram:

1. Set up the equation to calculate the missing height. No need to solve.
2. Let's say we redraw the histogram with different bins but with heights still measured in cm. One of the bars has a height of 0.0078 units of proportion per cm. If you re-measured the heights in meters instead of centimeters, what is the new height of the bar?