

<http://fathom.concord.org/resources/>
<http://fathom.concord.org/resources/tutorials/exploring-data-census-at-schools/>

FATHOM: Part 1 Exploring Data

Getting Data

Here are a few of the ways you can get data into Fathom.

1. Copy data from other programs (Word, Excel) and paste them into an empty collection. Choose **Import From File** from the **Import** submenu in the **File** menu. If importing doesn't work well, copy/paste might work better.
2. To access data sources from the Internet, simply drag the URLs from your Web browser into a Fathom document or by choosing **Import From URL** from the **Import** submenu in the **File** menu. However, some data will not import correctly or will require tweaking to format properly.
3. Of course you can simply type in data using the new data Table.
4. Open one of the many sample documents that come with Fathom. Choose **Open Sample Document** from the **File** menu and look in the **Sample Documents** folder.

Lab Data Description

We will use the data collected from 253 Texas State students who completed the survey during 2014/2015.

Open the **Students Survey Data** Fathom file.

On the **FATHOM** window double click on the **Collection 1** to open **Inspector**.

Select **Cases** and check the **Attributes**.

| Attributes (variables) | Survey's Questions |
|--------------------------|---|
| <i>Sex</i> | What is your sex? |
| <i>Texas</i> | Were you born in Texas? |
| <i>Hand</i> | What is your dominant hand? |
| <i>Eye color</i> | What is your eye color? |
| <i>Glasses</i> | Do you wear corrective lenses (glasses or contacts)? |
| <i>On campus</i> | Do you live on campus? |
| <i>Age</i> | How old are you in months? |
| <i>Height</i> | What is your height in inches? |
| <i>Weight</i> | What is your weight in pounds? |
| <i>Armspan</i> | What is your height in pounds? |
| <i>Foot length</i> | How many minutes did it take you to get from your home to your first class today? |
| <i>Hand length</i> | What was your mother's age when you were born? |
| <i>Snap Right</i> | What was your father's age when you were born? |
| <i>Snap Left</i> | What is the number of your siblings (including full, half and step)? |
| <i>Mother age</i> | What was your mother's age when you were born? |
| <i>Father age</i> | What was your father's age when you were born? |
| <i>Siblings</i> | How many siblings do you have? |
| <i>HS GPA</i> | What is your high school GPA? |
| <i>GPA</i> | What is your current GPA? |
| <i>Facebook</i> | What is the total number of your Facebook friends? |
| <i>Smoking</i> | How many cigarettes do you smoke per typical week? |
| <i>Drinking</i> | How many alcoholic beverages did you drink in the last 7 days? |
| <i>Sleep</i> | How many hours of sleep did you get last night? |
| <i>Messages sent</i> | How many text messages did you send yesterday? |
| <i>Messages received</i> | How many text messages did you receive yesterday? |

Important: To prevent unexpected changes in your data set, use the option **Prevent Changing Values in Graphs** from the menu **Collection**.

Categorical Data (One variable)

Describe one categorical variable:

1. Open a new **Graph**. Drag and drop attribute **Texas** from the inspector window.
2. Create a **Summary Table** for the variable **Texas**. Drag and drop **Summary** from the shelf. Drag and drop attribute **Texas** from the inspector window: frequencies appear.
3. Add relative frequencies (proportions) to the Summary table: click once on the Summary table and then use **Summary** from the main menu on the top of the Fathom screen. Select option **Add Formula** and type **columnproportion** in the formula window.

Categorical Data (Two variables)

Relationship between two categorical variables:

1. Create a new **Summary** for **Hand** and add an attribute **Glasses** to the top. Add **proportions** to the contingency table (*columnproportion* or *rowproportion*).
2. To illustrate graphically, use the segmented bar chart (convert Bar chart to **Ribbon** chart).

Numerical Data (One variable)

Describe the distribution of numerical variable, using graphs and numerical summaries.

1. The variable **Height** contains the student's heights measured in inches. Open a new **Graph** and drop the attribute Height on the x-axis. By default, you will see a dot plot.
2. From **Object** menu, choose **Duplicate Graph**.
Convert the Dotplot to a **Histogram**. Use an appropriate *binWidth*: double click on the Adjust the graph if needed.
3. Describe this distribution. Remember to mention shape, center, spread and outliers (if any).
4. Open the **Summary** for **Height**. The mean value appears automatically.
From the **Summary** menu, choose **Add Basic Statistics**.
From the **Summary** menu, choose **Add Five Number Summary**.
5. To add any statistic to the Summary table, use **Add Formula** from the menu **Summary**.
Use Function/Statistical/One Attribute.
6. Add a vertical line to plot the **Mean** directly on the histogram:
click once on the graph to select it; from the **Graph** menu, choose **Plot Value**. When the formula editor appears, type *mean(Height)*.) You can also select mean from Function/Statistical/One Attribute. (Same for the median etc.)
7. Add a horizontal line to plot: from the **Graph** menu, choose **Plot Function**. When the formula editor appears, type the numerical value you would like to show.
8. Duplicate your histogram and convert it to the **Boxplot**.
9. To compare two different histograms, convert them to the relative frequency histogram: click once on the graph to select it. Then, from the **Graph** menu, choose **Scale – Relative Frequency**.
10. To show two variables on the same histogram, add the second variable onto the vertical axes.
Example: Use the **relative frequency** histogram to show the distribution of **Heights** for male and female students on the same graph: add the variable **Sex** onto the vertical axis.
11. To delete any unused value from the Summary table: double click on it and delete from the formula window.

Numerical Data (Two variables)

Relationship between two numerical variables:

1. Select one pair of variables at your choice and determine their roles (dependent Y/independent X).
2. Open a new **Graph** and make a scatterplot. Drag both variables into the graph accordingly.
3. Describe the association (form, direction, strength).
4. To measure the strength of the linear relationship, find the **Correlation Coefficient**:
Open a new **Summary Table**, drag one attribute into the column head and another into the row head.
5. Create a Linear Model: go back to the scatterplot, select **Graph/Least-Squares Line**. The equation of linear regression model and the value of R^2 appear below the graph.
6. To observe the residuals, select **Graph/Make Residual Plot**.