



# FITTING A LINE TO DATA – EARNINGS AND EDUCATIONAL ATTAINMENT

## TEACHER VERSION

**Subject Level:**

Middle School Math

**Grade Level:**

8

**Approx. Time Required:**

45 minutes

**Learning Objectives:**

- Students will be able to create a scatter plot relating median annual income to years of education.
- Students will be able to informally fit a line on a scatter plot, derive an approximate equation for that line, and determine how well the line fits the data.

## Activity Description

Students will investigate the relationship between earnings and different levels of educational attainment by creating a scatter plot, adding an approximate line of best fit to the plot, and judging the line's accuracy.

---

**Suggested Grade Level:**

8

**Approximate Time Required:**

45 minutes

---

**Learning Objectives:**

- Students will be able to create a scatter plot relating median annual income to years of education.
- Students will be able to informally fit a line on a scatter plot, derive an approximate equation for that line, and determine how well the line fits the data.

---

**Topics:**

- Earnings
- Educational attainment
- Line of best fit
- Median annual income
- Scatter plots

**Skills Taught:**

- Creating a scatter plot
  - Determining the equation of a line of best fit
  - Informally fitting a line to data
-

## Materials Required

- The student version of this activity, 10 pages
- Rulers

A graphing calculator, graphing software (e.g., Microsoft Excel), or other graphing technology is optional.

## Activity Items

The following items are part of this activity. Items, their data sources, and any relevant instructions for viewing the source data online appear at the end of this teacher version.

- Item 1: Educational Attainment and Median Annual Income
- Item 2: Educational Attainment and Median Annual Income by Sex

For more information to help you introduce your students to the U.S. Census Bureau, read "[Census Bureau 101 for Students](#)." This information sheet can be printed and passed out to your students as well.

## Standards Addressed

See charts below. For more information, read

"[Overview of Education Standards and Guidelines Addressed in Statistics in Schools Activities](#)."

## Common Core State Standards for Mathematics

Standard	Domain	Cluster
<p><b>CCSS.MATH.CONTENT.8.SP.A.1</b></p> <p>Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.</p>	8 SP – Statistics & Probability	Investigate patterns of association in bivariate data.
<p><b>CCSS.MATH.CONTENT.8.SP.A.2</b></p> <p>Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.</p>	8 SP – Statistics & Probability	Investigate patterns of association in bivariate data.

## Common Core State Standards for Mathematical Practice

### Standard

**CCSS.MATH.PRACTICE.MP4. Model with mathematics.**

Students will create a scatter plot to model data and informally determine and assess the accuracy of a line of best fit. Students will also make predictions based on a linear model.

**CCSS.MATH.PRACTICE.MP5. Use appropriate tools strategically.**

Students will use a graph template to create a scatter plot, fit a line to that plot, and determine and interpret the line’s equation. Students could also use graphing technology to find the slope of the line of best fit for a second set of data.

## National Council of Teachers of Mathematics’ Principles and Standards for School Mathematics

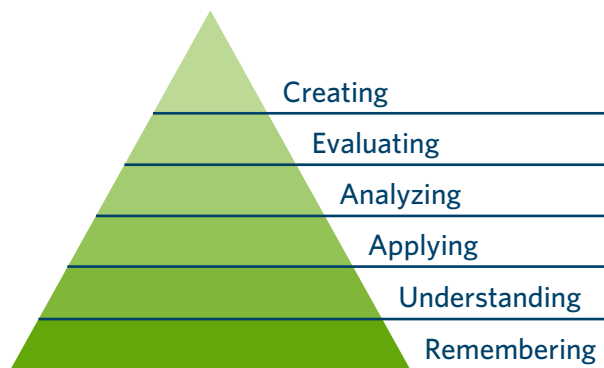
Content Standard	Students should be able to:	Expectation for Grade Band
Data Analysis and Probability	Select and use appropriate statistical methods to analyze data.	Discuss and understand the correspondence between data sets and their graphical representations, especially histograms, stem-and-leaf plots, box plots, and scatterplots.
Data Analysis and Probability	Develop and evaluate inferences and predictions that are based on data.	Make conjectures about possible relationships between two characteristics of a sample on the basis of scatterplots of the data and approximate lines of fit.

## Guidelines for Assessment and Instruction in Statistics Education

GAISE	Level A	Level B	Level C
Formulate Questions		X	
Collect Data			
Analyze Data		X	
Interpret Results		X	

## Bloom's Taxonomy

Students will **apply** their knowledge of **creating** scatter plots and writing linear equations to determine the association between two quantitative variables.



## Teacher Notes

### Before the Activity

Students must understand the following key terms:

- **Bivariate data** – pairs of linked numerical observations (e.g., a list of heights and weights for each player on a football team)
- **Scatter plot** – a graph in the coordinate plane that displays a set of bivariate data and can be used to determine how two variables are associated (e.g., to show associations between the heights and weights of players on a football team)
- **Line of best fit** – a straight line drawn through the center of a group of data points on a scatter plot, showing how closely the two variables on the scatter plot are associated
- **Slope** – the rate of change in a linear model, or the amount by which a y value increases (for positive slopes) or decreases (for negative slopes) for every unit increase in x value
- **Educational attainment** – the highest level of education that a person has completed
- **Confounding variable** – an outside variable that correlates with both the dependent and independent variables and could affect the data distribution

Students should have the following skills:

- Ability to identify associations of data and outliers on scatter plots
- Ability to write a line of best fit in slope-intercept form

Teachers should review with students how to create a scatter plot, informally fit a straight line to the scatter plot, and find the equation for that line — all by hand (without a calculator). Teachers should explain to students that the independent (explanatory) variable goes on the horizontal axis and the dependent (response) variable goes on the vertical axis.

### During the Activity

Teachers should monitor students as they work.

Teachers should be aware that the regression model used in this activity assumes the data distribution is linear; however, a nonlinear model might be a better fit.

To simplify this activity, teachers could create the scatter plots for students, requiring that students only fit the line and determine the equation. Teachers could also replace the written component of part 3 with the group discussion proposed for after the activity.

## After the Activity

Teachers could have students share their one-paragraph summaries from part 3 and could lead a class discussion on various ways that these data could be used in real-life situations. Teachers could also have students discuss the careers they would like to pursue and each career's required education level, directing students to refer back to the data to see the median income for each level — overall, and for men and women.

## Extension Idea

- Teachers could have more advanced students examine how a nonlinear model might fit these data, using graphing technology to determine both the  $r$  and  $r$ -squared values and then interpreting those values.

## Student Activity

Click [here](#) to download a printable version for students.

### Activity Items

The following items are part of this activity and appear at the end of this student version.

- Item 1: Educational Attainment and Median Annual Income
- Item 2: Educational Attainment and Median Annual Income by Sex

### Student Learning Objectives

- I will be able to create a scatter plot relating median annual income to years of education.
- I will be able to informally fit a line on a scatter plot, derive an approximate equation for that line, and determine how well the line fits the data.

What do you want to be when you grow up? How much money do you think you will make? How many years of education will you need for this career? These are important questions to ask yourself as you get older — surely you've heard them before from teachers or guardians. In this activity, you will explore data that could help you answer them.

The U.S. Census Bureau conducts a survey every month called the American Community Survey to collect data from a sample of the American population on things like income and education. **Item 1: Educational Attainment and Median Annual Income** shows the results from 2019 for workers 25 years and older.



Part 1

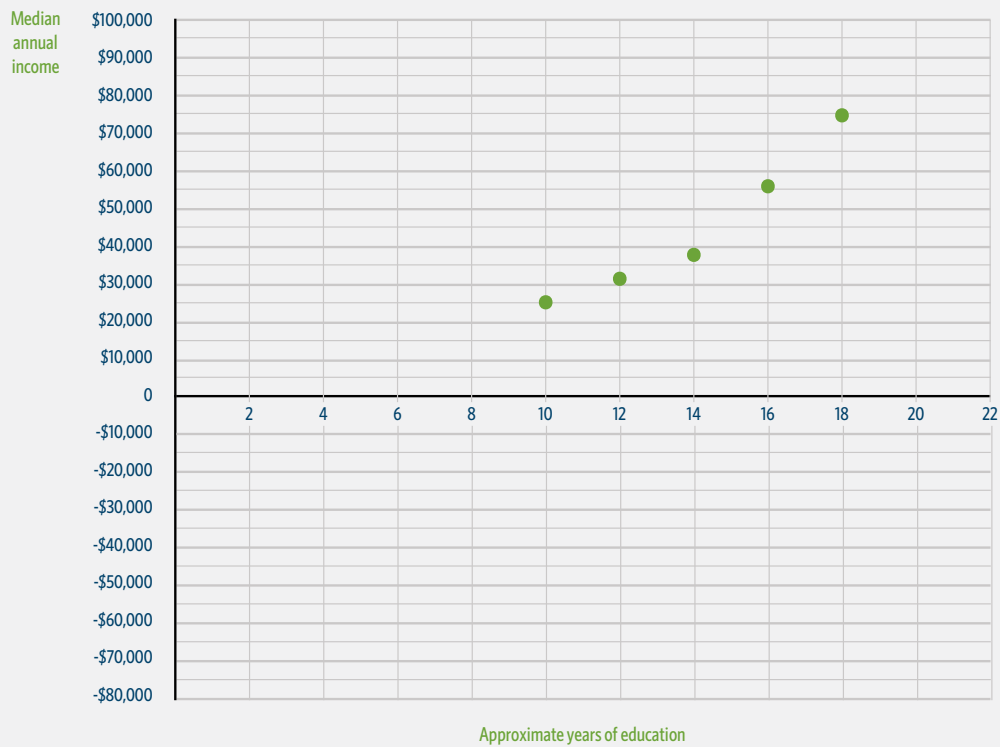
1. What patterns do you see in the data in **Item 1**?

**Student answers may vary but should mention that higher income is associated with more years of education.**

2. Create a scatter plot of these data on the graph template below, plotting the points as accurately as you can. Be sure to title your plot and label your axes.

**Student scatter plots should look similar to:**

Median Annual Income Based on Approximate Years of Education



3. Describe what you see in your scatter plot, keeping these questions in mind: Is the association positive, negative, or nonexistent? Do there appear to be any outliers or clusters of data? Are there any confounding variables that could be affecting the association?

**Student answers will vary, but should include that the association is generally positive and that there are no data points for educational attainment of less than 10 years (which makes sense because these data are about people 25 and older, and most U.S. schools require schooling through age 16).**

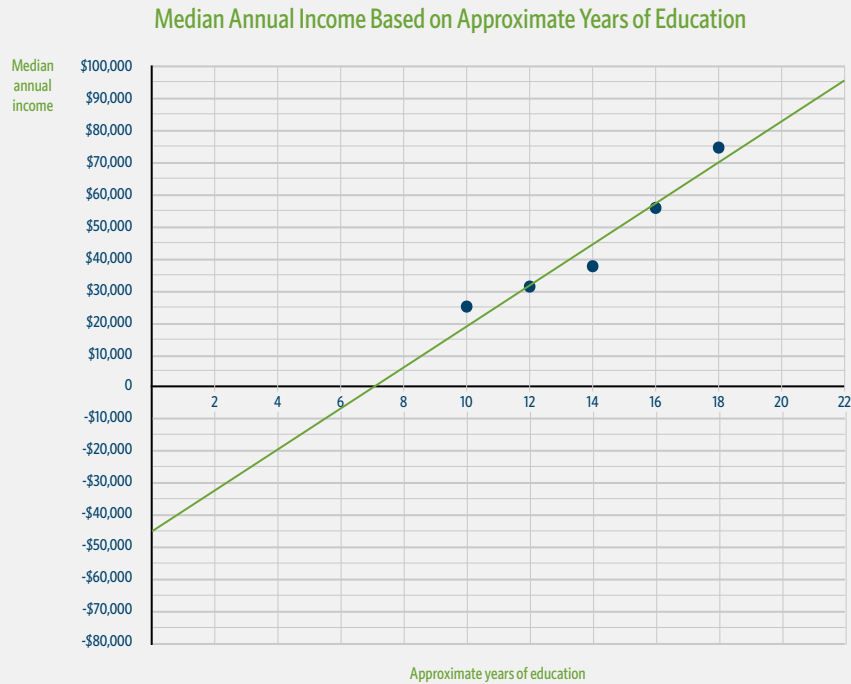
**There do not appear to be any outliers or clusters of data, which could be because the data set has only one point for each level of educational attainment.**

**The scatter plot shows a significant increase in median annual income between 14 and 16 years of education and between 16 and 18 years. These sharp increases could be the result of earning a four-year degree after high school and a master's degree after college, respectively.**

**Confounding variables like age and socioeconomic status could be affecting the association. For example, people with more years of education might also be older, meaning they may have more experience in the workforce, which usually leads to higher median annual incomes. Also, those with higher socioeconomic statuses might have more, or easier, access to better-paying jobs and to higher education.**

4. Next, use a ruler to draw a line on your scatter plot that best fits the data, extending it all the way to the vertical axis. This line should follow the general path and slope of your data points, with about the same number of points above and below it.

**Student lines should look similar to:**



5. Find a linear equation for your line.

**Student answers below will vary but should be similar to:**

- a. Locate the approximate y-intercept: -\$45,000
- b. Find the slope between any two points on your line (which may not necessarily be points from the original data!), rounding to the nearest thousand: 6,000. Show your work below.

**Student work will vary depending on the points chosen. For the points (8, 6,000) and (14, 44,000), for example, the slope is calculated as:  $(44,000 - 6,000) / (14 - 8) = 6,333.33$ . Rounded to the nearest thousand, that's 6,000.**

- c. Write the equation for the line in slope-intercept form (using the rounded slope).

**$y = 6,000x - 45,000$**

6. Is your y-intercept positive or negative? What does its value show in terms of earnings and educational attainment, and is that value useful to know in this case?

**The y-intercept is negative. In this case, the y-intercept indicates that a person with zero years of education will have a median annual income of about -\$45,000, which is not useful to know because it is impossible to have a negative median income and most people in the United States have at least some level of education.**

7. What does the slope tell us about earnings and educational attainment, and is that finding useful?

**The slope tells us that for each additional year of education people attain, their median annual income rises by about \$6,000, which is useful to know when estimating the monetary value of additional years of schooling.**

8. How well does your line fit the data? (Determine whether most or all of your points fall on, or are close to, the line.)

**The line fits well, as it goes through multiple data points and has a similar number of points above and below it.**

9. Turn to your classmates and compare your lines and equations. Do some of your classmates’ lines seem to better fit the data? Explain.

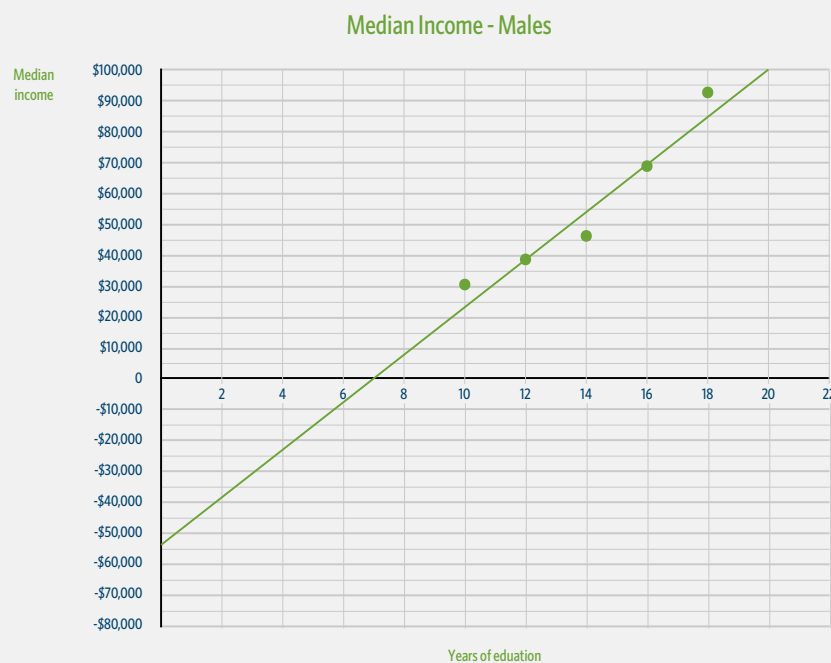
**Student answers will vary.**

## Part 2

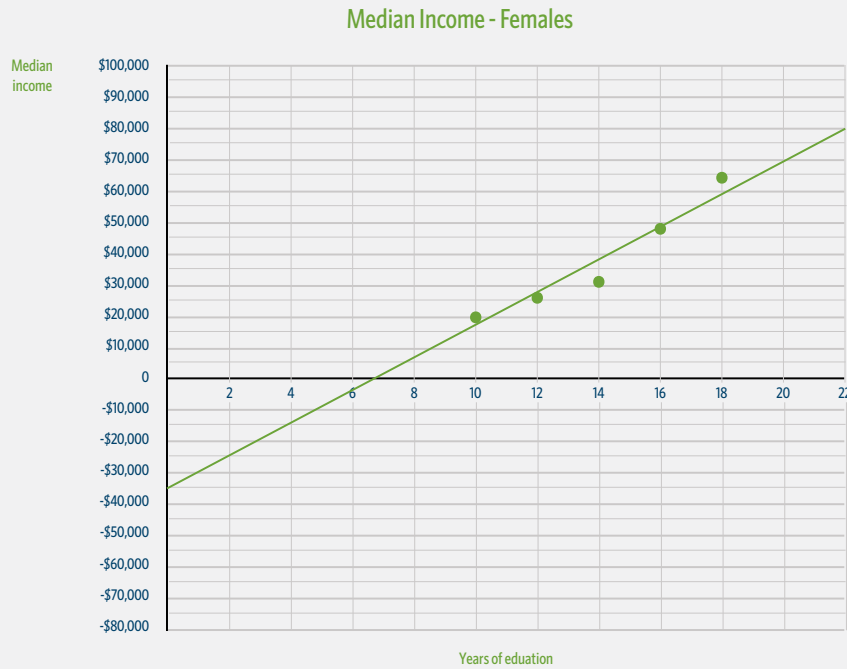
Now we will look at the educational attainment and median income data for men and for women.

Use the data in **Item 2: Educational Attainment and Median Annual Income by Sex** to create two scatter plots with lines of best fit — one for men and one for women (following the same process you used in part 1 to plot your points, draw your line, and determine your equation). Round all numbers to the nearest whole number.

**Student scatter plots and lines should look similar to:**



For men:  $y = \underline{8,000x - 53,000}$



For women:  $y = \underline{\underline{5,000x - 35,000}}$

- Which sex appears to have a higher rate of increase in earnings per year of education? How do you know?  
**Men. The increase in income per year of education (the slope) is nearly \$3,000 more for men than for women.**
- How does the slope of the line for men compare with the slope of the line from part 1?  
**The slope of the line for men is steeper than the slope of the first data set.**
- How does the slope of the line for women compare with the slope of the line from part 1?  
**The slope of the line for women is not as steep as the slope of the first data set.**

### Part 3

Write a one-paragraph summary, with a title, to describe what you learned about the relationship between earnings and educational attainment through creating your scatter plots, finding your lines of best fit, and determining their equations.

**Student summaries will vary, but possible titles include:**

- **Do Years Spent in College Pay Off?**
- **Do Women’s Levels of Education Affect Their Incomes?**

## Item 1: Educational Attainment and Median Annual Income

<b>Educational Attainment</b>	<b>Approximate Years of Education</b>	<b>Median Annual Income</b>
Less than high school graduate	10	\$25,876
High school graduate or equivalent	12	\$31,956
Some college or associate degree	14	\$38,125
Bachelor's degree	16	\$56,344
Graduate or professional degree	18	\$75,495

[data.census.gov/cedsci/table?q=B20004%3A%20MEDIAN%20EARNINGS%20IN%20THE%20PAST%2012%20MONTHS%20%28IN%202019%20INFLATION-ADJUSTED%20DOLLARS%29%20BY%20SEX%20BY%20EDUCATIONAL%20ATTAINMENT%20FOR%20THE%20POPULATION%2025%20YEARS%20AND%20OVER&tid=ACSDT1Y2019.B20004](https://data.census.gov/cedsci/table?q=B20004%3A%20MEDIAN%20EARNINGS%20IN%20THE%20PAST%2012%20MONTHS%20%28IN%202019%20INFLATION-ADJUSTED%20DOLLARS%29%20BY%20SEX%20BY%20EDUCATIONAL%20ATTAINMENT%20FOR%20THE%20POPULATION%2025%20YEARS%20AND%20OVER&tid=ACSDT1Y2019.B20004)

Copy and paste the link above into your browser to view the source data.

## Item 2: Educational Attainment and Median Annual Income by Sex

Men:		
Educational Attainment	Approximate Years of Education	Median Annual Income
Less than high school graduate	10	\$30,725
High school graduate or equivalent	12	\$38,906
Some college or associate degree	14	\$46,610
Bachelor's degree	16	\$69,201
Graduate or professional degree	18	\$93,998

Women:		
Educational Attainment	Approximate Years of Education	Median Annual Income
Less than high school graduate	10	\$20,046
High school graduate or equivalent	12	\$25,829
Some college or associate degree	14	\$31,644
Bachelor's degree	16	\$47,895
Graduate or professional degree	18	\$63,912

[data.census.gov/cedsci/table?q=B20004%3A%20MEDIAN%20EARNINGS%20IN%20THE%20PAST%2012%20MONTHS%20%28IN%202019%20INFLATION-ADJUSTED%20DOLLARS%29%20BY%20SEX%20BY%20EDUCATIONAL%20ATTAINMENT%20FOR%20THE%20POPULATION%2025%20YEARS%20AND%20OVER&tid=ACSDT1Y2019.B20004](https://data.census.gov/cedsci/table?q=B20004%3A%20MEDIAN%20EARNINGS%20IN%20THE%20PAST%2012%20MONTHS%20%28IN%202019%20INFLATION-ADJUSTED%20DOLLARS%29%20BY%20SEX%20BY%20EDUCATIONAL%20ATTAINMENT%20FOR%20THE%20POPULATION%2025%20YEARS%20AND%20OVER&tid=ACSDT1Y2019.B20004)

Copy and paste the link above into your browser to view the source data.