EllipSeeIt: Visualizing Strength and Direction of Correlation

Miguel Olvera
Cosumnes River College

Rafael E. Diaz
California State University Sacramento
rdiaz@csus.edu

Marco Molinaro, Matthew Steinwachs, Amy Smith, Tobias Spencer, Alberto Guzman, Alyssa Okita
University of California Davis, iAMSTEM Hub

STatistics Education Web
STEW
online journal of K-12 statistics lesson plans

Published: February, 2017

Overview of Lesson

This lesson focuses on correlation as a way of measuring the strength and direction of a linear association between two numerical variables. After reviewing some concepts of bivariate datasets, their graphical representations through scatterplots, and the meaning of a linear association, students will be asked to collect their own dataset by measuring two numerical variables about themselves. Then students enter their data online into the freely available statistical software *SeeIt* and produce scatterplots in order to determine if the association between two variables is linear (and therefore, whether it is appropriate to compute a correlation coefficient for each of these pairs of variables). Students use the dynamic feature of *SeeIt* to approximate the strength and direction of a linear association with an ellipse, and thus measure the correlation between two variables. Extensions include comparing strengths of correlation, the Pearson correlation coefficient, and how to fit a median-median line and a least-squares line when modeling a linear relationship between two variables.

GAISE Components

This investigation follows the four components of statistical problem solving put forth in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report. The four components are: formulate a question, design and implement a plan to collect data, analyze the data by measures and graphs, and interpret the results in the context of the original question. This is a **GAISE Level B** activity.

Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
5. Use appropriate tools strategically.

Learning Objectives Alignment with Common Core and NCTM PSSM

Learning Objectives	Common Core State Standards	NCTM Principles and Standards for School Mathematics
Students will be able to understand the concept of bivariate data sets by measuring and collecting pairs of quantitative variables about themselves. Students will learn how to graphically represent bivariate quantitative data sets using scatterplots. Students will learn how to visualize the direction and approximate the strength of linear association (correlation) without using a line as frame of reference.	8.SP.A.1. 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. HSS.ID.B6. Represent data on two quantitative variables on a scatterplot, and describe how the variables are related. HSS-.D.C8. Compute (using technology) and interpret the correlation coefficient of a linear fit.	Data Analysis and Probability Standards for Grades 9-12: **Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them:** • Understand the meaning of measurement data of Bivariate Data. **Select and use appropriate statistical methods to analyze data:** • Identify trends in Bivariate Data

Prerequisites
Students should have the ability to take measurements about themselves as well as construct a dataset table. Students should know how to construct a scatterplot. Students should have knowledge of how to use computer spreadsheets (e.g. Excel, Google Sheets).

Time Required
100 minutes, or two 50-minute class periods (first period gets longer as the number of students in class increases; therefore the first period may need to be finished during the second period, which tends to be shorter if students are well organized with computer facilities).

Materials and Preparation Required
- Computer with Internet access (Firefox, Safari, or Chrome only)
- *SeeIt* online tool for visualizing and analyzing bivariate quantitative data. Available at http://centerforbiophotonics.github.io/SeeIt3/correlations-stew.html
- Microsoft Office Excel or another spreadsheet program (e.g. Google Sheets)
- Printable Meter Ruler (link provided in the extension of this lesson plan, if needed)

EllipSeeIt: Visualizing Strength and Direction of Correlation
Teacher's Lesson Plan

Instructional Lesson Plan for Day 1

Introduction

Cobb et al (2003) argue that when students are introduced to the concept of correlation they should start by learning how to read scatterplots "vertically" instead of "diagonally." That is, students should start their instruction of correlation by visualizing scatterplots as the graphical representation of two sets of measurements that have been plotted orthogonally instead of visualizing the distances of the data points from the line of best fit. Therefore this lesson plan also emphasizes the visualization of the distribution of the y-values across the x-values when analyzing a scatterplot.

Begin the lesson by asking students to think of instances when they have heard that two things are correlated. Some examples are:

A) Income is correlated with years of education

B) Text messaging in class is correlated with low grades

Explain to students that statistical correlation is about measuring the strength of a linear association between two *numerical* or quantitative variables obtained from the *same* set of people, animals, plants, things, etc. One of these variables is called the independent variable (usually plotted on the *x*-axis), and the other is called the dependent variable (usually plotted on the *y*-axis). In statistical modeling the independent variable is called the explanatory variable, which is the one that may have an effect on, or explain, the dependent variable, which is called the response variable. Then elaborate on these topics for each of the above examples as follows:

For example A, tell students that researchers must have selected a group of people from whom they recorded the number of years of education they had after high school, and their income in say, thousands of dollars per year. It makes sense to think that the number of years of education may explain the annual salary of a person, and therefore to label salary as the explanatory variable (independent variable on the *x*-axis) of the response variable annual salary (dependent variable on the *y*-axis). Table 1 shows the first rows of a hypothetical dataset these researchers may have collected, and Figure 1 shows the scatterplot of this bivariate dataset where years of education are shown on the *x*–axis and salaries on the *y*-axis.

Name	Years of Education after High School (Independent or Explanatory Variable)	Annual Income ($1000s per year) (Dependent or Response Variable)
Joshua	2	40

Anne	5	68
Blake	0	22
. .	. . .	. . .

Table 1. First three rows of a hypothetical bivariate data for the number of years of education after high school and annual Income ($1000s per year) of a group of people.

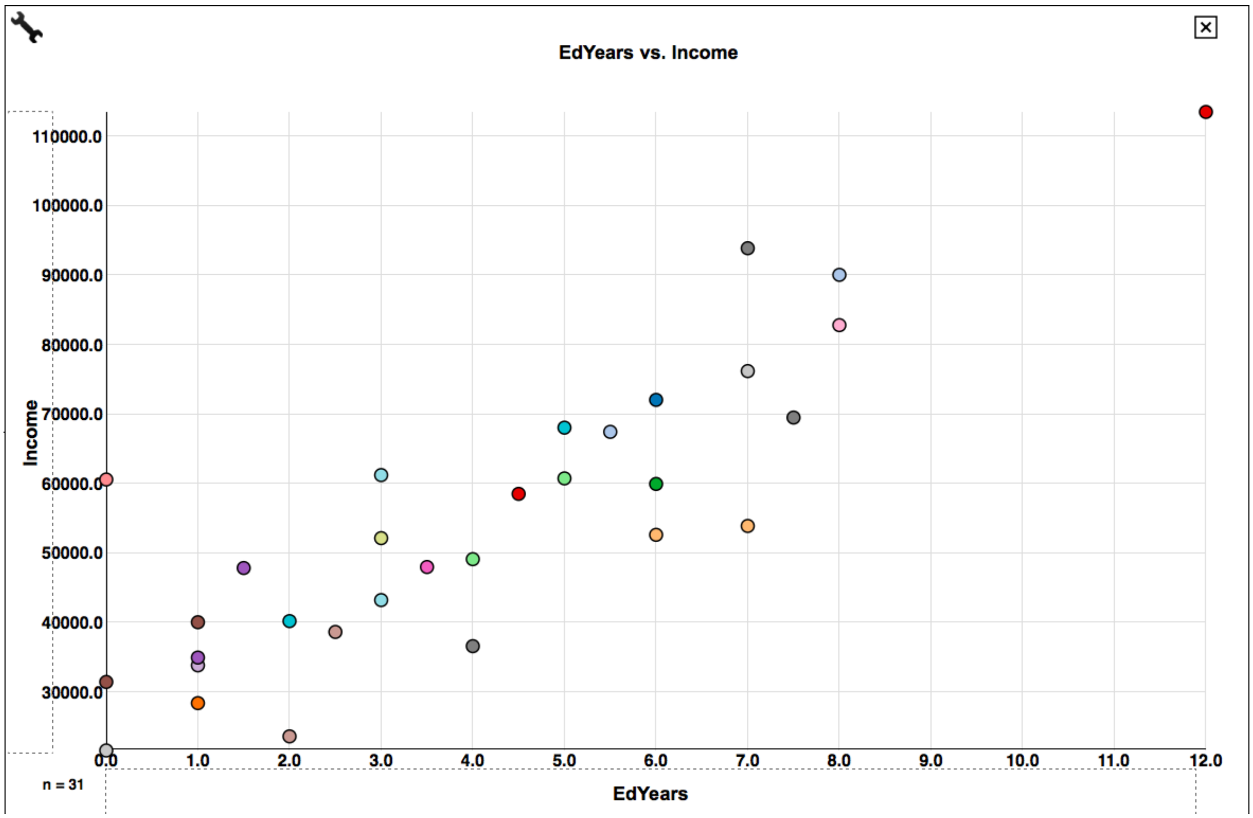


Figure 1. Scatterplot of annual income in $1000s/year (*y*-axis) versus number of years of education after high school (*x*-axis) for the hypothetical dataset described in Table 1.

In this lesson, we are using the free online tool *SeeIt*, designed specifically to visualize and analyze bivariate quantitative data. *SeeIt* assigns colors (at random) to the data points in the scatterplots to help in the identification of individual points during discussions.

Ask students to answer Question 1 on the handout for day one.

Students should be able to conclude that the association between these variables appears linear. That is that the ranges of the response variable in general shift upwards, at about a constant rate, as the values of the explanatory variable increase (such as the y-values in a line with *positive* slope in this case). Therefore, it is appropriate to measure the direction and strength of association (correlation) between these two variables. Tell students that in this case the *direction* of the association between years of education and salary is *positive* since the values salary in general increase, at about a constant rate, as the values of years of education increase.

For example B, tell students that researchers must have selected a group of students from whom they recorded, say, the number of text messages they sent while in class during the two weeks prior to an exam, and their *percentage* grade (0 to 100) on the exam. It makes sense to think that the number of text messages may explain the percentage grade on the exam, and therefore label the number of text messages as the explanatory variable (independent variable on the *x*-axis) of the response variable percentage grade (dependent variable on the *y*-axis). Table 2 shows the first rows of a hypothetical dataset these researchers may have collected, and Figure 2 shows the scatterplot of this bivariate dataset where number of text messages are shown on the *x*–axis and percentage grades on the *y*-axis.

Student	Number of Text Messages Sent In Class (Independent or Explanatory Variable)	Grade on Exam (Percentage) (Dependent or Response Variable)
Lucas	0	95
Amanda	21	74
Michael	40	68
. .	. .	. .

Table 2. First three rows of a hypothetical bivariate data for the number of text messages sent during a class for a period of two weeks prior to an exam, and the percentage grade on the exam of a group of students.

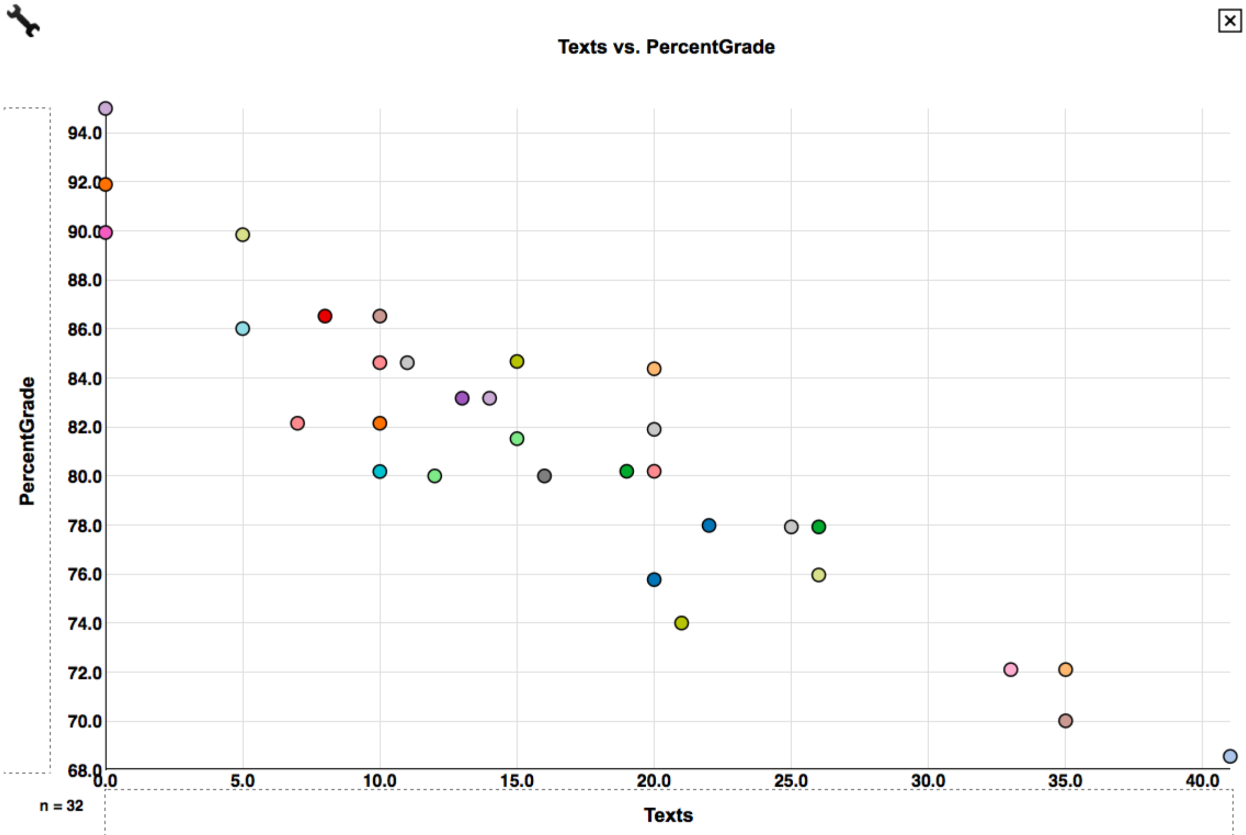


Figure 2. Scatterplot of the number of text messages sent during class over a two-week period prior to an exam, and the percentage grade in the exam for the hypothetical dataset described in Table 3.

Ask students to answer Question 2 on the handout for day one.

Students should be able to conclude that the association of these variables is linear. That is that the ranges of the response variable in general shift downwards, at about a constant rate, as the values of the explanatory variable increase (such as the y-values in a line with *negative* slope in this case). Therefore, it is appropriate to measure the direction and strength of association (correlation) between these two variables. Tell students that in this case the *direction* of the association between text messages and percentage grade is *negative* since the values of percentage grade in general decrease, at about a constant rate, as the number of text messages increase.

It is important that students understand that a correlation cannot be measured when the scatterplot shows an association other than linear. Show students scatterplot in Figure 3 as an example where it is not appropriate to measure the correlation. This scatterplot shows a hypothetical dataset where a group of students were asked to study a subject for up to an hour, and then take a test designed to measure the students’ retention of the material of this subject as a percentage. The graph indicates that as the number of minutes of study goes up, so does the percentage retention score in the test as long as the number of minutes goes from zero to 30 minutes. After 30 minutes and up to 60 minutes, the graph indicates that students decreased their retention score the longer they studied continuously the material. The relationship between number of minutes studied and retention percentage scores is rather parabolic instead of linear throughout the interval 0 to 60 minutes. Therefore, it does not make sense to obtain a measure of the direction and strength of a linear relationship (correlation) in this case. Another statistical technique would be needed to describe the parabolic relationship between these variables.

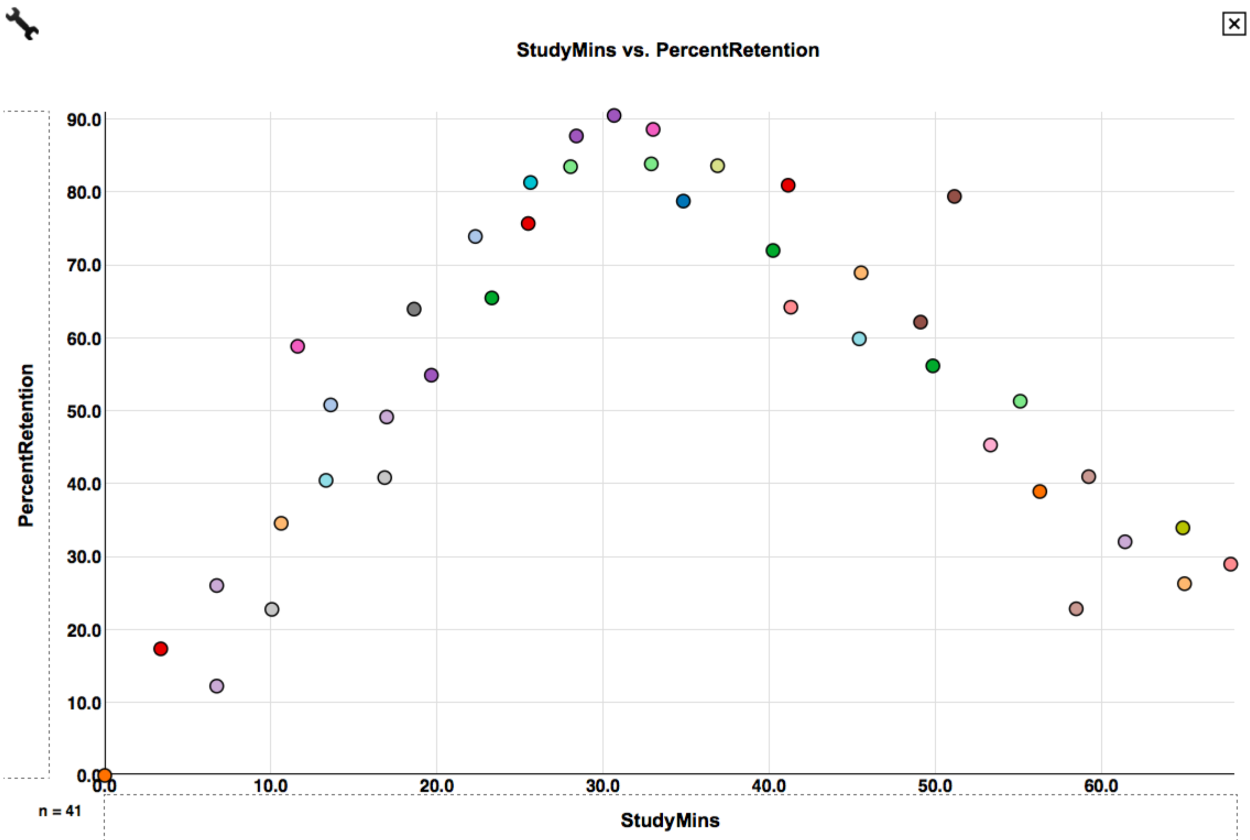


Figure 3. Scatterplot of minutes of study and percent retention.

Finally, show students an example of a scatterplot that shows that sometimes two numerical variables may not show any kind of association. In this instance, the ranges of the response variable vary randomly throughout the values of the explanatory variable. Figure 4 below shows that there is no line or curve that could be used to describe the association between the number of hours spent on Facebook per day and the number of friends in this social media of a hypothetical group of 31 students. In this case it is still valid to talk about the correlation between these two variables; however, the correlation is said to be none or zero.

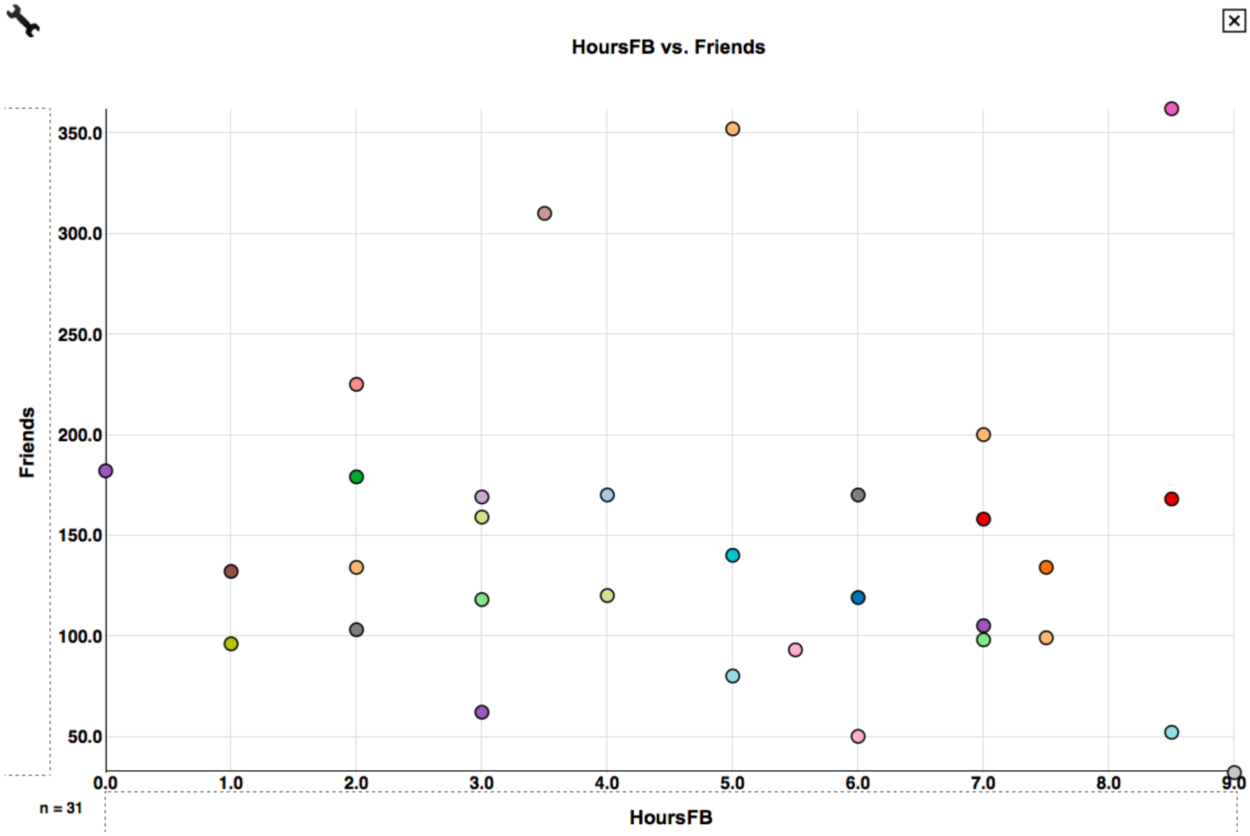


Figure 4. Scatterplot of number of hours spent on Facebook per day and number of friends in this social media.

I. Formulate a Question

Ask students to think of pairs of variables about themselves that they can measure in class to explore their correlation. Examples of variables students may suggest to measure and analyze pairwise are height, weight, armspan, text messages sent during a period of time, number of hours slept a previous night and, say, numerical grades on an exam the following day, etc. (an extension of this lesson plan offers support for measuring and comparing the correlation between height versus armspan and height versus right foot length).

Once students have selected a pair of variables to explore, ask them to label one of them as the independent or explanatory variable and the other as the dependent or response variable. Note that for some pairs of variable, this labeling may be interchangeable. Then ask students if they think the pair of variables they are suggesting to explore are linearly associated, and therefore if it is valid to measure their correlation. That is, ask students, as the values of explanatory variable increase/decrease, do the values of the response variable increase/decrease at about a steady rate?

II. Design and Implement a Plan to Collect Data

Tell students that they will collect their own dataset by measuring the two variables they have chosen to analyze.

Students can input their data set directly into the software program *SeeIt* (see appendix). However, if students don't have access to a computer in the classroom, the teacher can record the data on a spreadsheet (Excel or Google Sheet) in his/her computer, and provide it to the students to analyze in *SeeIt* once students have access to a computer.

Once students have access to a computer, ask them to go into *SeeIt* at:
http://centerforbiophotonics.github.io/SeeIt3/correlations-stew.html

Then follow the instructions in the appendix to create a data set in *SeeIt*. This short appendix concentrates on inputting a data set into *SeeIt*; however, to explore all features of *SeeIt*, see http://centerforbiophotonics.github.io/SeeIt3/disthelp.html)

Note: once data is entered into *SeeIt*, you can visualize and analyze it. However, you can not save the data in *SeeIt*, thus if you close the web browser and reopen *SeeIt*, your entered data will not be there. Thus, it is important to save your original data in a spreadsheet (e.g. Excel or Google Sheet).

III. Analyze the Data

Students will start analyzing their data by learning how to create and visualize scatterplots in *SeeIt*. Ask students to click on the title of their dataset in the dataset panel so that the names of the variables they measured appear below the title of their data set. To create a scatterplot ask students to click on the name of the variable they are using as the explanatory variable and drag it into the empty rectangle on the x-axis of the graphing window of *SeeIt*. Then click and drag the name of the response variable drag it into the empty rectangle on the y-axis.

Ask students to analyze the scatterplots by answering Question 3 on the activity sheet.

IV. Interpret the Data

For the students in your class, are the values of the response and explanatory variables approximately linearly associated? If so, what is the direction of the linear association? Justify your answer.

Finish this section by telling students that on the second day of the lesson they will learn about the strength of a linear association (correlation), and how to approximate it using a special feature of *SeeIt*.

Suggested Assessment Day 1

The data in the following table corresponds to the number of fast food meals that a group of individuals ate the week prior to measuring their body mass index (BMI). Answer the questions below using this bivariate dataset.

Name	Mean Fast Food Meals per Week	BMI
Lucas	0	17.49
Rob	3	25
Marie	2	19.04
Diana	2	22
Isabel	1	18
Tracy	4	26
Roberto	5	28
Lyn	2	20.58
Danielle	5	29
Blake	4	23
Troy	7	32
Peach	1	19
Leonel	3	23.53
Randy	3	22.41
Laura	3	21.15
Cassandra	4	24.8
Vladimir	6	29.96
Gregory	4	23.78
Michelle	3	24.31
Bianca	2	19.95

a. Construct a scatterplot for this bivariate dataset.

b. For people who ate 2 fast food meals, their BMI ranges from _____ to _____.

c. For people who ate 4 fast food meals, their BMI ranges from _____ to _____.

d. In general, what happens to the ranges of BMI as the average number of fast food meals eaten the prior week increases?

e. Is there an approximate linear association between BMI and number of fast food meals eaten the prior week? Is so, what is the direction of this association?

Instructional lesson plan for Day 2

I. Formulate Question

Start by introducing student to the concept of direction of a linear association. Remind students that they have already observed on day one examples of datasets where, in general, the values of the dependent or response variable either go up (e.g. fast food meals versus BMI) or go down (number of text messages versus percentage grade in an exam) as the values of the independent or explanatory variable increase. When the values of the response variable go up as the values of the explanatory variable go up it is said that the correlation or linear association between two variables is *positive*; whereas when the values of the response variable go down as the values of the explanatory variable go up it is said that the correlation is *negative*.

Pose the following question:
How strong is the association between our two variables? Can we develop a way to express this strength as a measure?

Show students that they can informally measure the strength of a correlation as *none*, *medium* and *strong* for both positive and negative directions by using the metaphors of elliptical shapes in Figure 5.

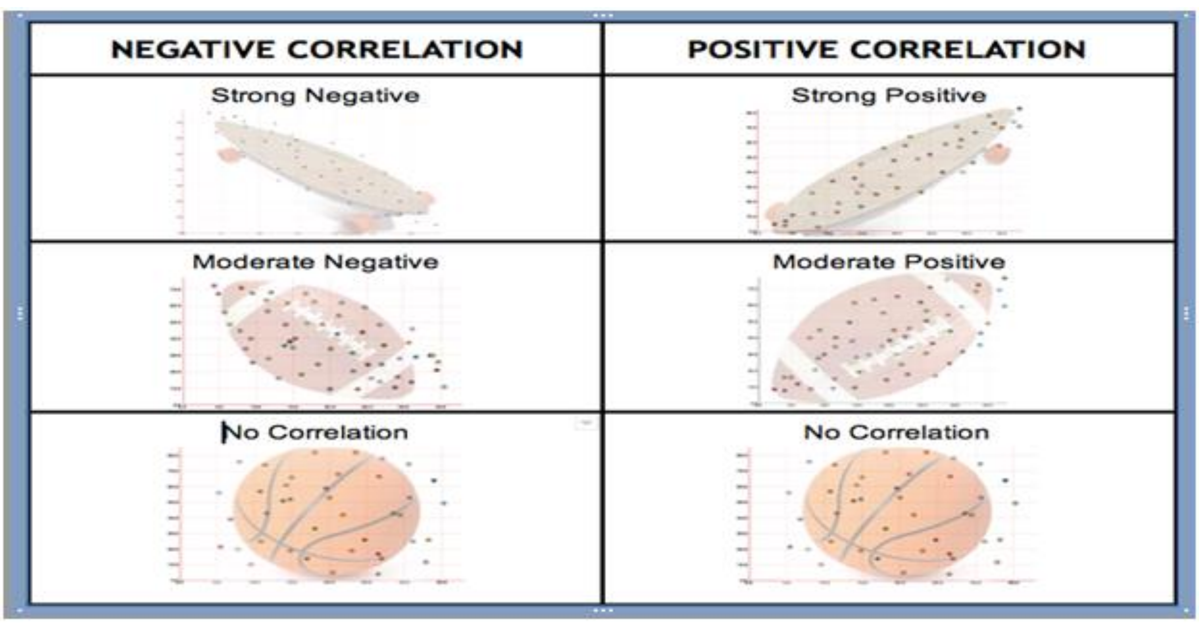


Figure 5. Metaphors of elliptical shapes illustrating the strength of a correlation as strong, medium and low or none in both positive and negative directions.

Figure 5 shows that a strong correlation usually results in a scatterplot where an ellipse that encloses nearly all of the data points resembles a skateboard. Note also that a strong correlation can be either positive or negative (first row in Figure 5). Similarly, a medium correlation usually results in an ellipse that can resemble a football, (second row in Figure 5) and a low or no correlation results in an ellipse that resembles a basketball (third row in Figure 5). In more general terms, the tighter the ellipse that encloses nearly all of the points in of a scatterplot, the stronger the correlation.

II. Collect Data

Students will be using the data they collected on day one to investigate the strength and direction of the correlation between pairs of variables.

They need to re-load their dataset (saved on day one on a spreadsheet, e.g. Excel or Google Sheet) into *SeeIt*, and reproduce the scatterplot they created in day one.

III. Analyze the Data

Ask students to imagine ellipses that would enclose nearly all of the points in the scatterplot, and chose one of the shapes in Figure 5 to describe the strength and direction of the correlation. They will use a special feature of *SeeIt* that allows them to draw an ellipse enclosing nearly all of the points in a scatterplot in order for them to visualize and approximate the strength and direction of correlation.

Click on the wrench icon in the upper-left corner of the scatterplot of height versus armspan. A dropdown window will appear, where students can click on the box next to "show ellipse." Close the dropdown window by clicking on the X on its upper-right corner. A red ellipse will appear at the center of the scatterplot as shown in the Figure 6.

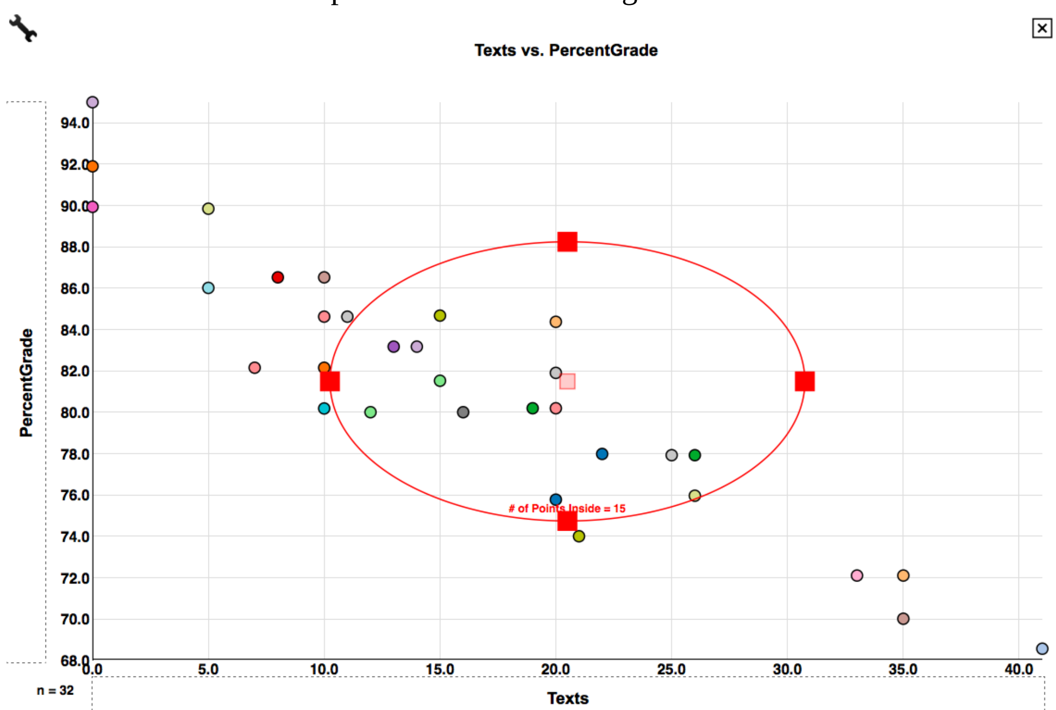


Figure 6. Ellipse produced by *SeeIt* that can be adjusted to enclose most points in a scatterplot.

Ask students to place the cursor on the red squares on the ellipse to adjust the ellipse so that it encloses nearly all of the points in the scatterplot. The pink square in the center of the ellipse can be used to move the center of the ellipse. After doing this, ask students to repeat the above steps to enclose, in an ellipse, nearly all of the points in the scatterplot of their data.

IV. Interpret the Data

Ask students to describe the strength and direction of the correlation of their data and, and answer Question 1 of the handout for day two.

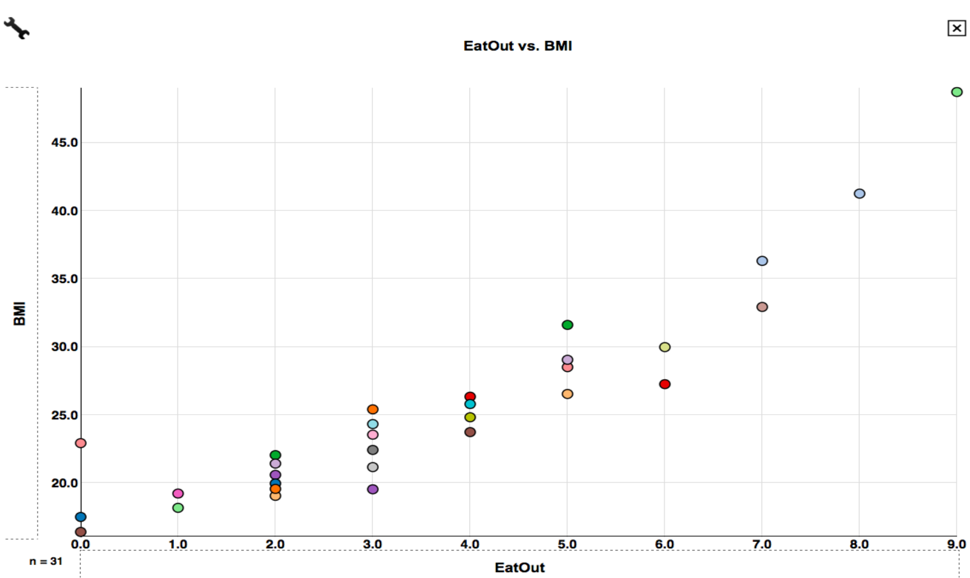
Now students will deepen their knowledge about the strength of correlation by performing an activity that will allow them to compare and interpret the strength of correlation between two scatterplots. Tell students that they will create side by side scatterplots of the datasets TVHoursVsGPA (*x*-axis: HoursTV, *y*-axis: GPA) TextingVsPercentGrade (*x*-axis: Texts, *y*-axis: PercentGrade) in *SeeIt*. To do this, click on the "Two Graphs" button located on the top bar of the graphing window, and then click and drag the variables as described above. Now ask student to complete question 2 of the handout

Suggested Assessment Day 2

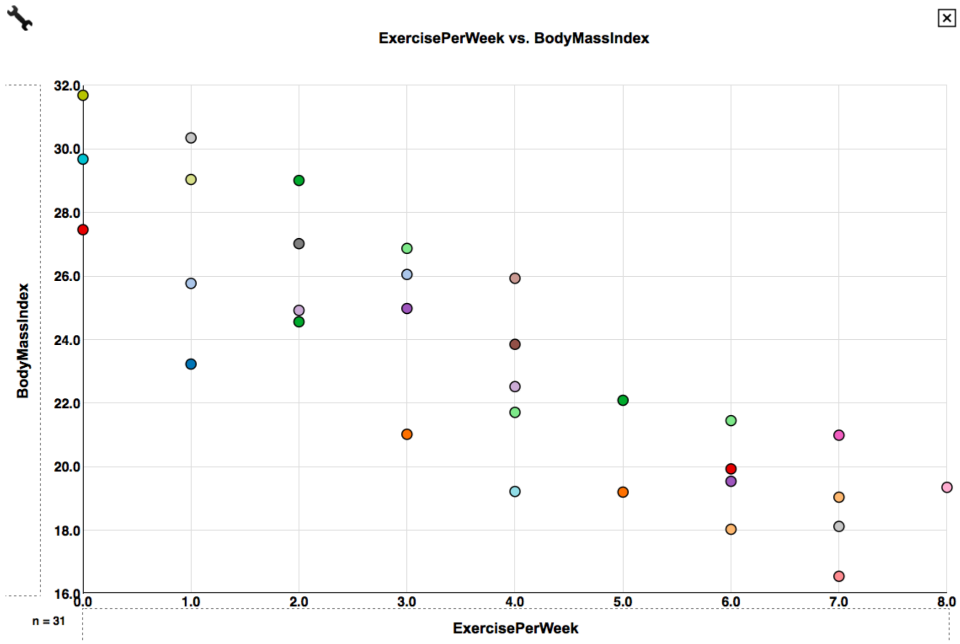
For each of the following plots:

1) Draw an ellipse to enclose most of the points of the following scatterplots;
2) Identify the direction (positive or negative) of the association;
3) Describe the strength of the association between the variables involved.

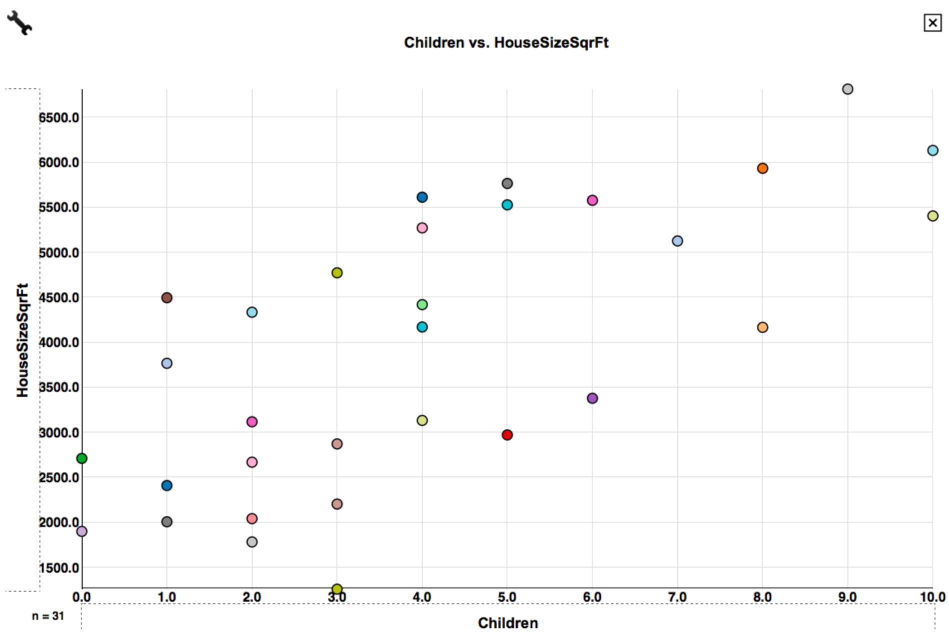
a. X- axis: Average number of fast-food meals a group of people eat per week.
Y-axis: Body Mass Index (BMI) of each of these people.



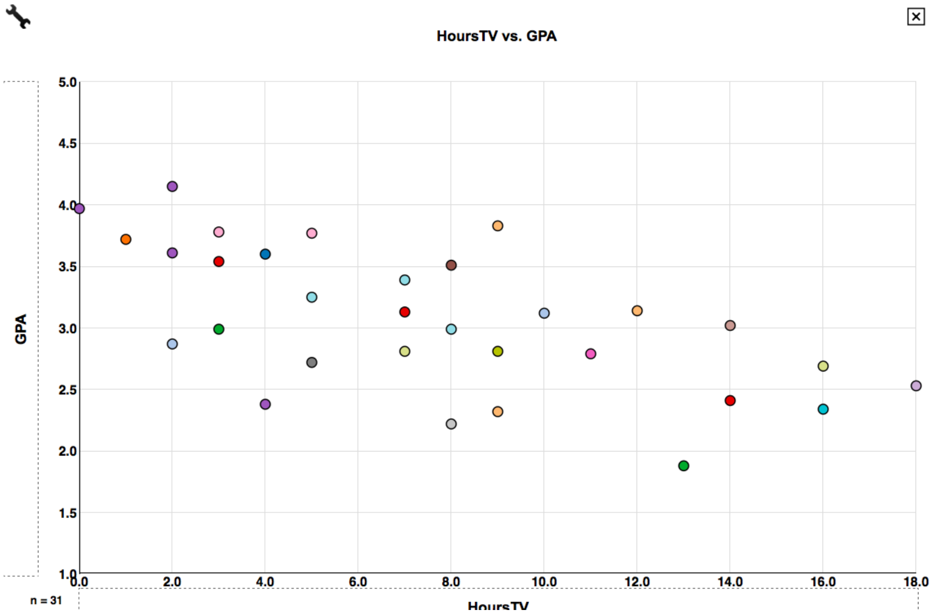
b. X- axis: Average number of days a group of people exercise physically per week
   Y-axis: Body Mass Index (BMI) of each of these people.



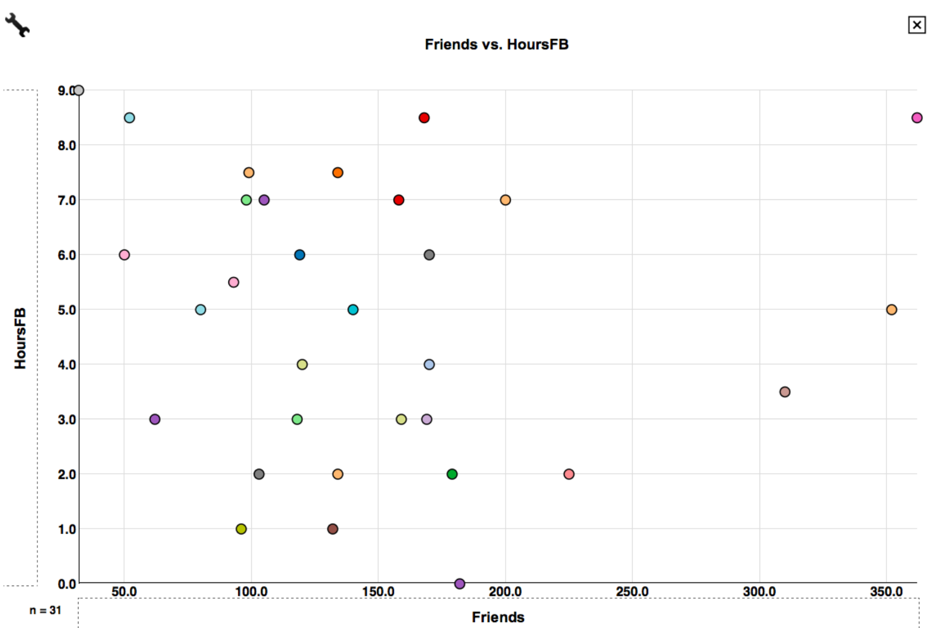
c. X- axis: Number of children per family for a group of families.
   Y-axis: Square footage of the homes of these families.



d. X- axis: Average number of hours of TV a group of high school students watches per week.
Y-axis: GPA of each student.



e. X- axis: Average number of hours spent on Facebook by a group of high school students.
Y-axis: Number of Facebook friends each student has.



Possible Extensions

Analyzing and Comparing the Correlation between Armspan and Height and between Height and Foot Length.

Analysis of a large dataset (http://www.censusatschool.ie/get-data/results/phase-10-1011, Jan, 21, 2016), of more than 10,000 students from Ireland who completed a survey from Census at School (an international classroom project that engages students in grades 4 – 12 in statistical problem solving - for the U.S. see https://www.amstat.org/censusatschool/index.cfm), showed there was a stronger correlation between armspan and height than between height and foot length. This extension hopes that the data from your class would reproduce this result in order to introduce how to interpret and compare the difference between two correlation strengths.

The data collection process can be conducted by taping a vertical and a horizontal centimeter ruler to a wall for students to measure their height and armspan. You can download a 2-meter printable ruler here:

Students can streamline the input of the data of the entire class directly into the software program *SeeIt* (see appendix) as each student steps up to the rulers on the wall and reads his/her measurements out loud to the rest of the class. Ask students to measure their right foot in centimeters with a regular ruler prior to stepping up to the rulers on the wall so they also can give this measurement to the rest of the class when they read their height and armspan.

Linear Modeling

SeeIt can also fit median-median and least-squares regression lines using special visualization tools, and it can also compute the Pearson correlation coefficient. You can extend the teaching of this lesson to include these extra features, which can be accessed under the wrench icon on the upper-left corner in the graph of a scatterplot: click on the wrench icon located on the upper-left corner of the graphing window and check the boxes under **Least squares** that say Show line?, Show squares, Show equation, Show r value? for all features of a least squares line (Figure 7 below); or check the boxes under **Median-median** that say Show divisions, Show dots, Show line, Show equation? for all features of a median-median line (see e.g. http://www.mathedpage.org/attc/lessons/ch.12/12.02-median-median-line.pdf) (Figure 8 below).

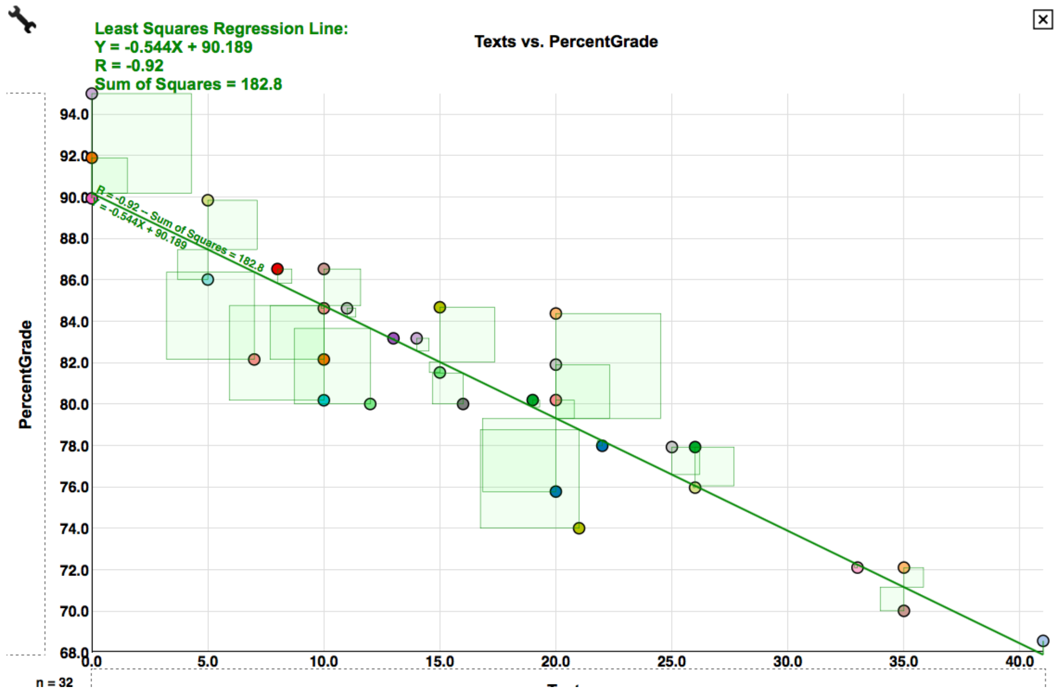


Figure 7. Least squares line, areas of squares with minimum total area, equation of line and correlation coefficient r.

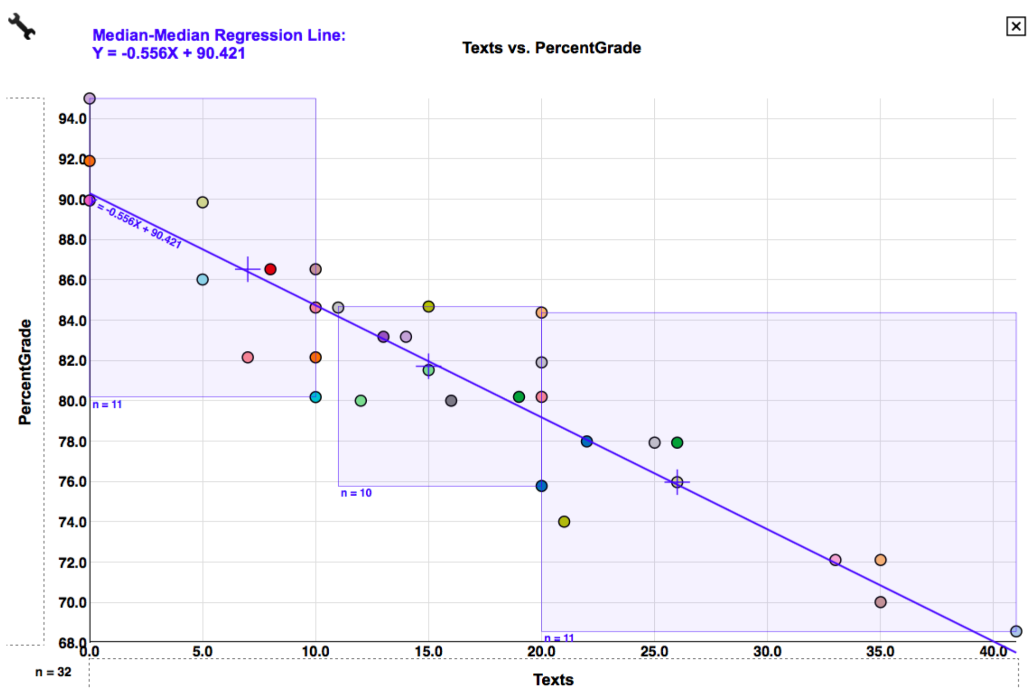


Figure 8. Median-median line and its equation.

References

1. Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction*, 21(1), 1–78.

2. Reeves, S.L, Varakamin, C. Henry, C.J. (1996). The relationship between arm-span measurement and height with special reference to gender and ethnicity. *European Journal of Clinical Nutrition*, 50(6), 398-400.

3. *SeeIt* [Software] (2011). Davis California: University of California Davis, iAMSTEM Hub. n.d. Web. 15 July 2014. <https://sites.google.com/a/cbst.ucdavis.edu/sbcepublic/SeeIt>.

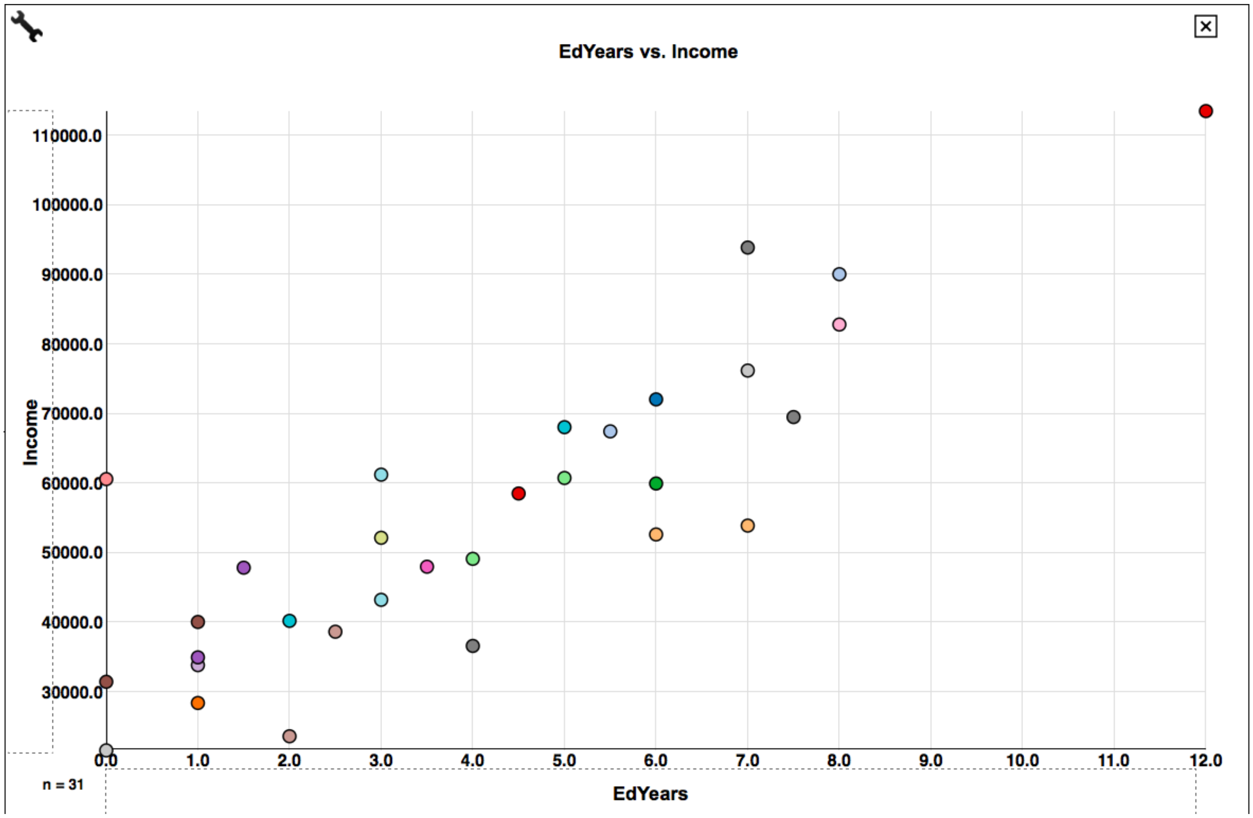
4. The *CensusAtSchool* Project run by the International Centre for Statistical Education.

Acknowledgments

We thank Matthew Steinwachs for his software programming skills in implementing the correlation visualization tools in *SeeIt*. We also thank Alisa Lee, Education Manager at the UC Davis iAMSTEM Hub, for her valuable help in managing the project "How Sure Are You? Science, Biostatistics and Cancer Education," which funded the educational research involved in several components of this lesson plan through an NIH Science Education Partnership Award (SEPA). Finally, we thank all participating high school teachers for all their feedback while pilot testing the materials used for this lesson plan.

EllipSeeIt: Visualizing Strength and Direction of Correlation
Student Handout Day 1

1. Answer the following questions by analyzing the following scatterplot .

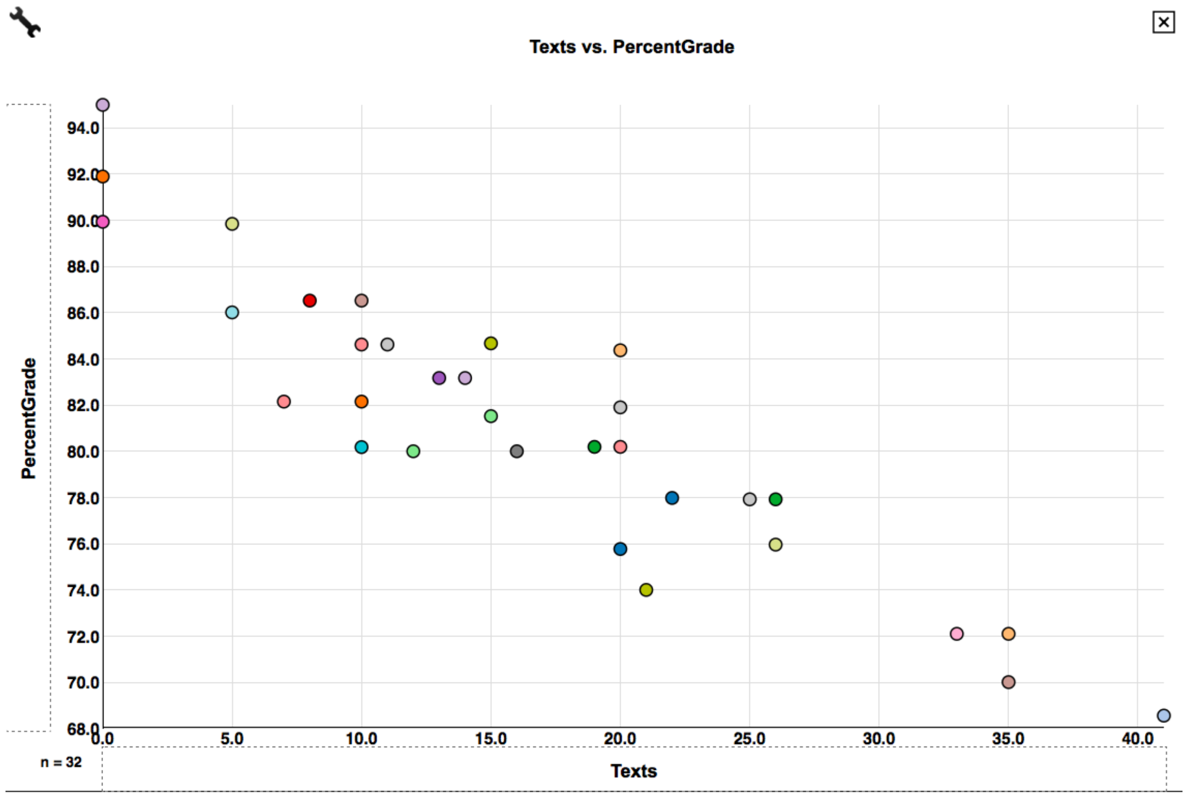


a. For people in this dataset who have 1 year of education after high school, annual income ranges from _______ to _______.

b. For people in this dataset who have 6 years of education after high school, annual income ranges from _______ to _______.

c. Repeat the instructions of parts a and b above scanning the whole range of the years of education. What happens to the ranges of annual income in general as the number of years of education after high school increases? Do these ranges tend to shift upward (as in a line with a positive slope) through the entire range of years of education? What can you say about the general trend between the annual income of a person and the number of years of education after high school?

2. Answer the following questions by analyzing the following scatterplot.



a. For students in this dataset who never sent a single text message during class, their percentage grade on the exam ranges from _______ to _______.

b. For students in this dataset who sent an average of 10 text messages during class, their percentage grade on the exam ranges from _______ to _______.

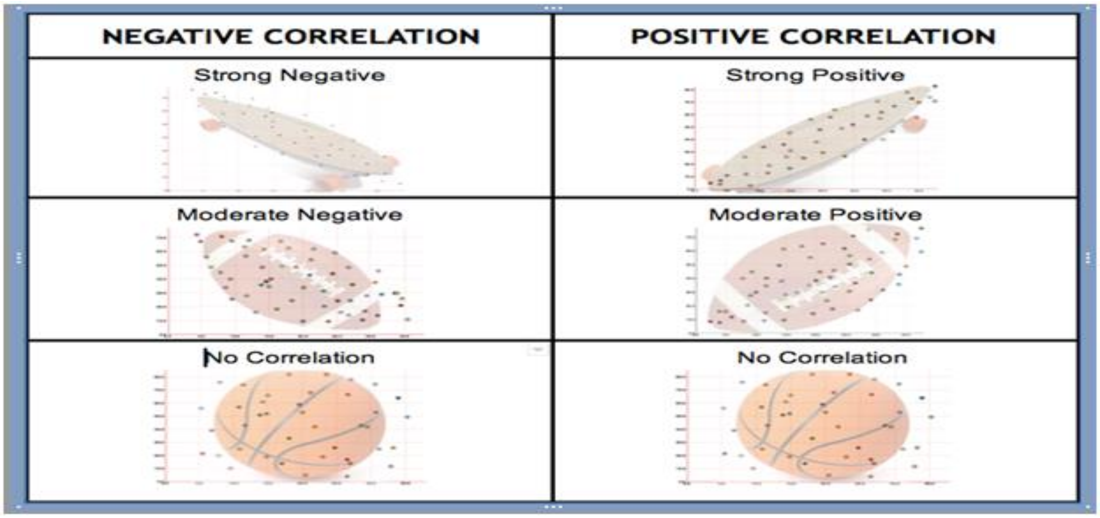
c. Repeat the instructions of parts a an b above scanning the whole range of average text messages sent during class. What happens to the ranges of grades in general as the number of text messages sent during class increases? Do these ranges tend to shift downward (as in a line with a negative slope) through the entire range of number of text messages? What can you say about the general trend between the number text messages sent in class and exam grades?

3. Answer the following question by analyzing the scatterplots of the pair of variables you chose to analyze.

What can you say about the ranges of the response variable as the values of explanatory variable increase? Is the relationship between these variables approximately linear?

EllipSeeIt: Visualizing Strength and Direction of Correlation
Student Handout Day 2

1. Using the shapes shown below, how would you describe the strength and direction of the correlation between the variables you chose to analyze?



2. Answer the following questions after creating scatterplots for the datasets TVHoursVsGPA (*x*-axis: HoursTV, *y*-axis: GPA) and TextingVsPercentGrade (*x*-axis: Texts, *y*-axis: PercentGrade) in *SeeIt* .

a. Is there an approximate linear relationship between GPA and the number daily hours of watched TV increases for this group of students? Explain.

b. Using the shapes shown in, how would you describe the strength and direction of the correlation between TVHours and GPA after enclosing most of the points in an ellipse?

c. Is there an approximate linear relationship between text messages sent during class and percentage grades on the exam for this group of students? Explain.

d. Using the shapes, how would you describe the strength and direction of the correlation between Texts and PercentGrade after enclosing most of the points in an ellipse?

e. Is the strength of the correlation of TVHours versus GPA stronger, weaker or equal to that of Texts versus PercentGrade?

Student Handouts
Sample Solutions for the Teacher

Day 1 Handout

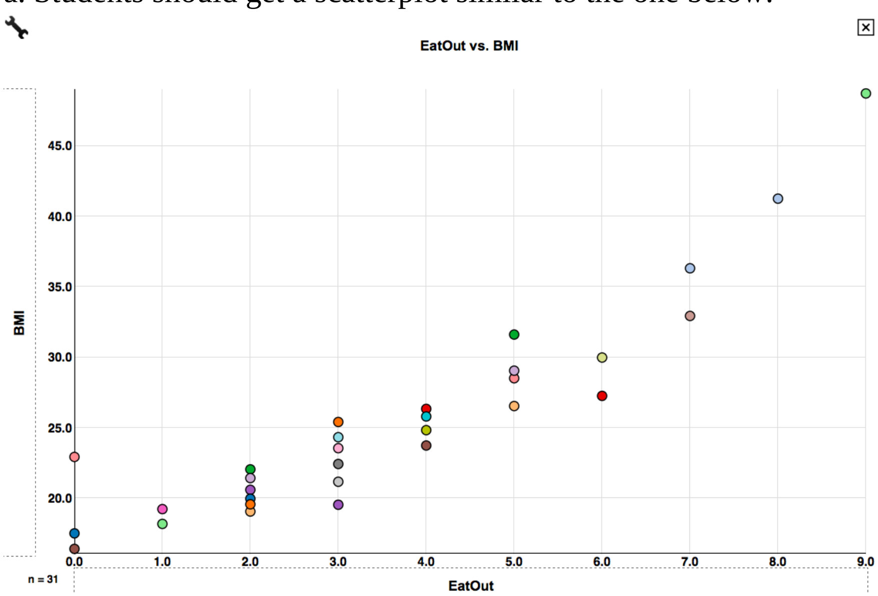
1. a. From about $28,000.00 to about $40,000.00.
   b. From about $52,000.00 to about $72,000.00.
   c. Even though the ranges of income vary and overlap among the years of education, in general these ranges tend to shift upward as the number of years of education increases. Therefore, in general income goes up as the years of education go up.

2. a. From about 90 percent to about 95 percent.
   b. From about 80 percent to about 86 percent.
   c. Even though the ranges of grades vary and overlap among the number of text messages sent, in general these ranges tend to shift downward as the number of text messages increases. Therefore, in general grades go down as number of text messages increases.

3. Answers vary depending on the relationship of the chosen variables: positive or negative linear, non-linear, or none.

Day 1 Suggested Assessment Items

a. Students should get a scatterplot similar to the one below:



b. From about 19 to 22.
c. From about 23 to 26.
d. As the average number of fast food meals increases the BMI generally increases as well, at about a constant rate.

e. Yes. The direction of the approximate linear association is positive.

Day 2 Handout

1. Answers vary: strong, moderate, none; positive, negative.
2. a. Yes. GPA decreases at about a steady rate as the number of hours of TV watch increases.

b. Football-shaped, i.e., moderate and negative.

c. Yes. Percentage grade decreases at about a steady rate as the number of text messages sent increases.

d. Scateboard-shaped, i.e., strong and negative.

e. The correlation between TVHours and GPA is weaker than the correlation between Texts and PercentGrade.

Day 2 Suggested Assessment

a. Skateboarded-shaped, i.e., strong positive;

b. Skateboarded-shaped, i.e., strong negative;

c. Football-shaped, i.e., medium positive;

d. Football-shaped, i.e., medium negative;

e. Basketball-shpaped, i.e., none.

Appendix
Creating a Data Set in *SeeIt*

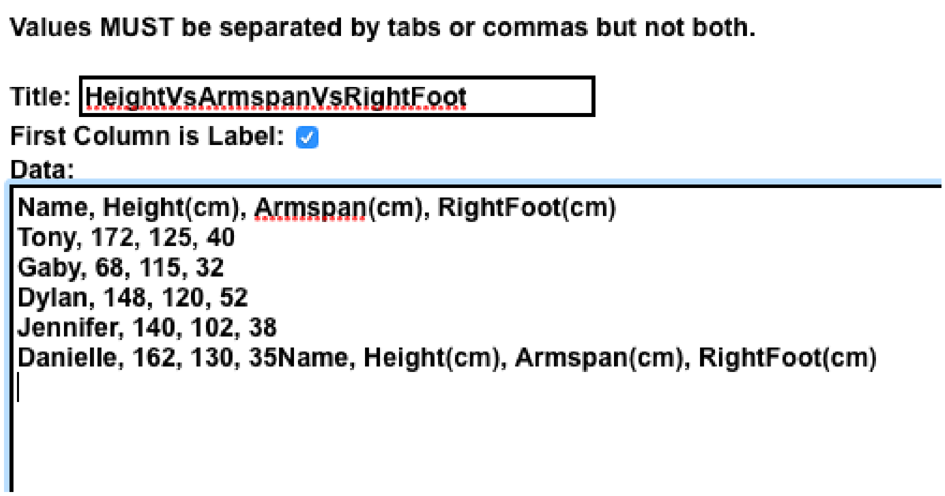
http://centerforbiophotonics.github.io/SeeIt3/correlations-stew.html

Once students have loaded the *SeeIt* tool using the link above, ask them to click on “Add a Worksheet” at the bottom of the dataset panel located on the left of the screen in order to display a *SeeIt* edit worksheet. Ask students to add a title to their dataset in the rectangle that says ***Enter Worksheet Title***. Students must follow the following format when entering the dataset into the worksheet:

a) The first row must have the labels (each one a single word) of the columns separated by a comma followed by a space, e.g., Name, Height(cm), Armspan(cm), RightFoot(cm)

b) All rows after the first one need to have the name of a student (one single word), and the measurements of the variables they have chosen to analyze separated by a comma followed by a space. Note that students may choose to measure more than two variables to be analyzed pairwise. Figure below illustrates the format needed for a data set from a group of students collecting measurements of their height, armspan and right foot length.

Note: if the teacher has already collected the students’ data on a spreadsheet (Excel or Google Sheet), he/she can email it to the students so that they can simply select and copy the data from the worksheet, and paste it directly onto the *SeeIt* edit worksheet. Note that when pasting the data from an Excel, or a Google Sheet, the columns of data are separated by tabs, which is another format of datasets accepted by *SeeIt*.



Format required to enter data onto an edit worksheet in *SeeIt*.

Once students have finished collecting their dataset, ask them to select the entire dataset in the *SeeIt* edit worksheet, copy it, and paste it into an Excel or Google Sheet. Then students can save this spreadsheet by emailing it to themselves as an attachment or storing it in a portable USB drive to use the second day of the lesson plan. Students will need to paste back their dataset into *SeeIt* the second day of the lesson plan since their data will be lost once they close *SeeIt* (datasets entered into *SeeIt* will appear again in the dataset panel of another *SeeIt* session only if the browser's cookies are not deleted in between *SeeIt* sessions).

After students are done saving their data, ask them to click on the "Load Worksheet From Form" button located at bottom of the *Seeit* edit worksheet. This will refresh the data set panel (left side of *SeeIt* window) where students will be able to see the title of their data set). Note that the dataset panel can be shown or not by clicking on the button that says "Show Datasets" or "Hide Datasets" located at the upper left corner of the graphing window of *SeeIt*.