3.1 Scatterplots and Correlation KFP AP Stats Lesson

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Do daily death counts correlate with air pollution?

Scatterplots and Correlation

**Description**

*This lesson plan utilizes candy, sports, and air pollution to explore scatterplots and correlation of variables. It is adapted from Section 3.1 in The Practice of Statistics Sixth Edition by Starnes and Tabor.*

**Packet Contents**

* [Introduction](#intro)
* [Curriculum Alignment](#curriculum)
* [Objectives](#objectives)
* [Time and Location](#timeloc)
* [Teacher Materials](#teachermat)
* [Student Materials](#studentmat)
* [Safety](#safety)
* [Student Prior Knowledge](#studentpriorknowledge)
* [Teacher Preparation](#teacherprep)
* [Activities](#activities)
* [Assessment](#assessment)
* [Critical Vocabulary](#vocab)
* [Author Information](#authorinfo)

**Introduction**

Students will build scatterplots representative of a variety of different data: candy versus hand span, diamonds versus the SAT, payroll versus wins, height versus amount of sleep, sugar versus calories, and deaths versus air pollution and heat. They will learn how to describe the plots: the explanatory and response variables, the direction and strength of the association, and how to find the correlation coefficient.

**Curriculum Alignment**

S-ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

**Objectives**

The objectives list what students are expected to learn after completing the lesson plan.

* Students will be able to distinguish between explanatory and response variables for quantitative data.
* Students will be able to make a scatterplot to display the relationship between two quantitative variables.
* Students will be able to describe the direction, form, and strength of a relationship displayed in a scatterplot and identify unusual features.
* Students will be able to interpret the correlation.

**Time & Location**

Classroom

One 90-minute period

**Teacher Materials**

**The Practice of Statistics Sixth Edition by Starnes and Tabor (can use other editions, but page numbers will be associated with this edition**

**Bowl full of wrapped candies**

**Whiteboard**

**Whiteboard markers**

**Projector**

**Computer**

**Doc Cam (for the discussions, showing student work, if desired)**

**Student Materials**

Graph paper

Notebook paper

Graphing Calculators

Pencils

Rulers

**Student Prior Knowledge**

**Students should be able to plot coordinate points on a graph.**

**Students should be able to choose appropriate scales for graph axes.**

**Teacher Preparations**

Teacher needs to have a bowl of wrapped candies out in a prominent location, and rulers need to be readily available to students.

On the board, teacher should have the instructions from the “Candy Grab” activity on page 152.

Teacher should also have a table with the headings “Hand span (cm)” and “Number of candies” written on the board beside the instructions, as shown in the activity.

Students can be grouped however you like; I tend to prefer groups of two.

Teacher should either have enough books for each group to use, enough photocopies for each group to use, a projection of the tables/scatterplots/diagrams for students to look at in class, or an accessible electronic copy for students to use.

Have the *Guess the Correlation* applet loaded ([www.rossmanchance.com/applets](http://www.rossmanchance.com/applets)) [activity described on p.161]

**Activities**

10 min: Candy Grab activity (p.152, the first page of Chapter 3)

7 min: Show <https://www.cbsnews.com/news/pollution-makes-air-in-parts-of-california-dangerous-to-breathe/>, provide <http://www.latimes.com/nation/la-na-utah-smog-2017-story.html>, provide <https://www.aging.ca.gov/data_and_statistics/facts_about_elderly/>, <https://blissair.com/what-is-pm-2-5.htm>, and tell them that we have a 13 year data set that shows us the daily death counts (all, all over 65, all between 65 and 74, all over 75, each of those options minus accidents, each of these options for circulatory deaths, each of these options for respiratory deaths), PM2.5 daily average (and the PM2.5davg from 1 day before, 2 days before, and 3 days before), Ozone daily average, minimum temperature, and maximum temperature.

5 min: Open up <https://padlet.com/mcleak/deathandpollution> (password: apstats) or create your own padlet and have students type in ANY curiosity, question, wonder, concern, worry, research, or learning desire that pops in their head. Encourage them to like other people’s post-its and comment on other people’s post-its.

5 min: Hand each group one of the South Coast data sets and have them create a scatterplot using that information

5 min: Take notes on explanatory and response variable (p.153); do this as you desire: could be the students read it from the book or handout, could be a PowerPoint projection, could be a discussion, could be written on the board, could be guided notes, however you like to give notes for your students. I would focus on what the words “explanatory” and “response” mean in their everyday lives so that they can make sense of the fact that the explanatory variable *explains* how the response variable is *responding*, and show them where those names came from in the first place.

2 min: Label explanatory and response variable for all of the scatterplots we’ve already done, plus the example on p. 154

5 min: Discuss as a class what the answers are to the explanatory and response variables (have someone from each group describe how they thought about it to the whole class, see if you can include different ideas or thought processes from when you were walking around). I advise using the random number generator on your calculator to pick who speaks, since that gets them used to randomization early on AND encourages them to work hard to be ready to answer, since it is perceived as fair, and it is unpredictable. (I have my tables in groups of four, so I generate numbers 1-4).

5 min: Students take notes on how to make a scatterplot (p.154-155). Make sure to point out here that statisticians only pay attention to the section of the scatterplot that is relevant to them, so the numbers on the axes could start anywhere.

5 min: Students take notes on how to describe a scatterplot (p.156-157). (form, direction, outliers, and strength)

5 min: Quick Application: Students describe the scatterplots we already made, describe the scatterplots on p.156 -p.158 (independently or as a group, while the teacher walks around the room).

5 min: Discuss how they were described (have someone from each group describe how they thought about it to the whole class, see if you can include different ideas or thought processes from when you were walking around).

10 min: p.159, technology corner. Teach students how to find a scatterplot on their calculators.

Ask:

* Why would we want to do this in a calculator?
* Are there limitations to what we can do in the calculator?
* Can you think of any other technologies that would be better to learn on for those limitations?

5 min: take basic correlation notes on p.160-161. It is important to note here that the correlation can only describe direction (negative/positive) and strength (weak, moderately weak, none, moderately strong, strong) of the linear relationship between the variables. The number can’t say “72% strength” or “72% of the data is…” or anything like that, unfortunately. This is tough for students at first, but they get used to it as they practice.

10 min: play *Guess the Correlation* as a whole class (activity instructions are on p.161), after each guess, have students describe why they guessed that number, and discuss important pieces to be looking for on the graph to help with what the correlation might be.

6 min: Exit Ticket (Check Your Understanding on p.162; add the instruction: label the explanatory and response variables; when grading, look for the description words they learned earlier) and have students write down the which variables they would be interested in exploring from the air pollution.

HW: Choose from Section 3.1 Exercises p.171-175

**Assessment**

Exit Ticket (Check Your Understanding on p.162; add the instruction: label the explanatory and response variables; when grading, look for the description words they learned earlier)

Assesses Objectives:

* Students will be able to distinguish between explanatory and response variables for quantitative data.
* Students will be able to describe the direction, form, and strength of a relationship displayed in a scatterplot and identify unusual features.
* Students will be able to interpret the correlation.

**Formative assessment (basically all of the activities)**

* **Specifically, the “make a scatterplot using these differing data sets from South Coast CA” assesses the “**Students will be able to make a scatterplot to display the relationship between two quantitative variables.”
* Suggested questions:
	+ How did you decide which variable was which?
	+ How could knowing the strength and direction be useful to a statistician?
	+ What forms would concern you and why?
	+ What did you consider when guessing the correlation coefficient?
	+ Were there any correlations that you were surprised about? Why do you think that occurred?

**Critical Vocabulary**

Response Variable: measures an outcome of a study

Explanatory Variable: may help predict or explain changes in a response variable

Scatterplot: shows the relationship between two quantitative variables measured on the same individuals.

Positive association: when the above-average values of one variable tend to accompany above-average values of the other variable and when below-average values also tend to occur together

Negative association: when above-average values of one variable tend to accompany below-average values of another variable

No association: if knowing the value of one variable does not help us predict the value of another variable

Direction: a scatterplot can show a positive association, a negative association, or no association

Form: a scatterplot can show a linear form (follows a straight line) or a nonlinear form

Strength: a scatterplot can show a weak, moderate or strong association (if the points don’t deviate much from the form identified)

Unusual features: outliers that fall outside the overall pattern and distinct clusters of points

Correlation (r): measures the direction and strength of a linear association between two quantitative variables

**Author Information**

**In this section, tell us about yourself and your mentor! Include the following:**

**Margaret Borden:**

* **Knightdale High School of Collaborative Design, Wake County Public School System, Raleigh**
* **9-12 Math**
* **Second year**
* mcleak@ncsu.edu**;** mborden@wcpss.net

**Richard Smith:**

* **SAMSI and UNC Chapel Hill Department of Statistics and Operations Research**
* **Statistical Research**
* **Mark L. Reed III Distinguished Professor, Director of SAMSI, http://www.unc.edu/~rls/**
* rls@samsi.info