**\*All solutions and teacher notes in blue\***

**AP Statistics Handout Key: Lesson 3.1**

*Topics:* explanatory/response, describing scatterplots, correlation coefficient (r), causation

**Lesson 3.1 Guided Notes**

*Explanatory and response variables*

Bivariate data: data with **two** variables. Two quantitative variables are visualized in a **scatterplot**

**Income and Food Access Example (H.E.B Grocery Stores) \***

|  |  |  |  |
| --- | --- | --- | --- |
| **Zip Code** | **Grocery Store Location** | **Average Household Income (x)** | **Organic Vegetables Offered (y)** |
| 78204 | South Flores | $71,186 | 36 |
| 78207 | N. Rosillo st | $34,234 | 4 |
| 78204 | Nogalitos st | $71,186 | 28 |
| 78201 | Frederickburg rd | $48,760 | 31 |
| 78212 | Olmos | $78,096 | 78 |
| 78202 | New Braunfels | $40,506 | 14 |
| 78237 | Castroville | $38,166 | 12 |
| 78228 | Culebra rd | $50,398 | 18 |
| 78227 | Marbach rd | $49,437 | 38 |
| 78240 | Babcock rd | $66,073 | 84 |
| 78230 | Wurzbach rd | $86,566 | 61 |
| 78251 | W Loop 1604 N | $78,176 | 56 |
| 78238 | Bandera rd | $59,154 | 62 |
| 78223 | S.New Braunfels | $50,252 | 44 |
| 78221 | SW Military | $48,364 | 26 |
| 78224 | S Zarzamora | $56,274 | 29 |
| 78220 | W.W. White rd | $41,318 | 15 |
| 78209 | East basse rd | $125,145 | 95 |
| 78216 | San pedro | $65,911 | 18 |
| 78223 | S.E Military dr | $50,252 | 65 |
| 78218 | Austiin hwy | $53,945 | 50 |
| 78213 | West Avenue | $59,072 | 35 |
| 78227 | Valley Hi dr | $49,437 | 36 |
| 78244 | Foster dr | $72,080 | 28 |
| 78231 | N.W Military | $108,486 | 95 |
| 78239 | Montogomery | $70,530 | 46 |
| 78217 | Perrinbeiter rd | $57,199 | 29 |
| 78251 | FM 471 west | $78,176 | 73 |
| 78250 | Guilbeau rd | $78,288 | 53 |
| 78230 | De Zavala | $86,566 | 86 |
| 78247 | Thousand oaks | $84,181 | 68 |
| 78247 | O’Connor rd | $84,181 | 56 |
| 78251 | Potranco rd | $78,176 | 85 |
| 78247 | Bulverde rd | $84,181 | 86 |
| 78248 | NW Loop 1604 | $135,547 | 93 |
| 78232 | 18140 San Pedro | $92,946 | 82 |
| 78249 | 9238 Loop 1004 | $77,894 | 96 |

Using this data from San Antonio, TX, we will explore whether there is a relationship between neighborhood income and access to organic items at local grocery stores.

Explanatory (independent) variable: the variable that predicts, explains, or influences a trend in the response variable. This is the **x-variable**.

Response (dependent) variable: the measured outcome. Responds to trends in the explanatory variable. This is the

**y-variable**.

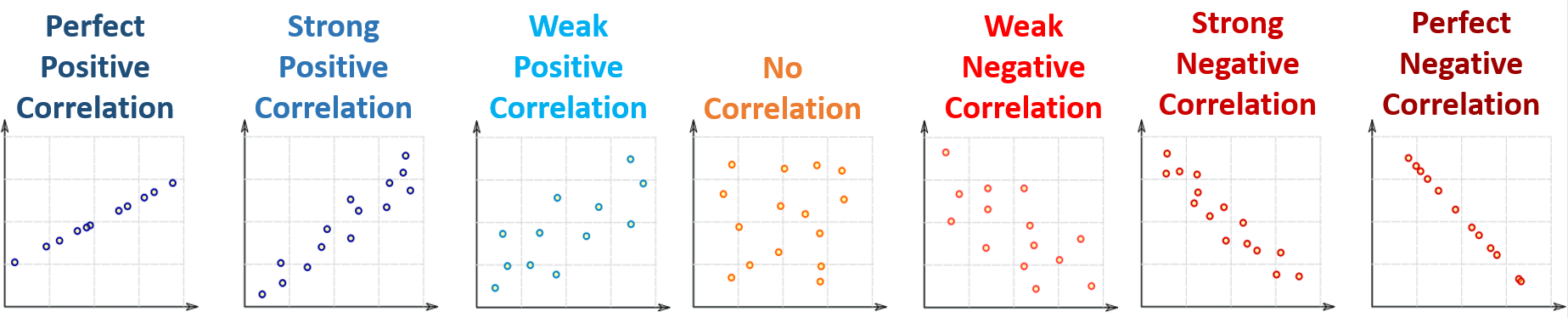
In this example, which variable is the explanatory variable? Why?

The explanatory variable is “average household income” because we are investigating how well it predicts or relates to the measured outcome: the number of organic vegetables offered at the local supermarket.

In this example, which variable is the response variable? Why?

The response variable is “organic vegetables offered” because it’s the “outcome” we’re measuring, based on average household income.

\*Dataset compiled by student Linda Saucedo, Fall 2019

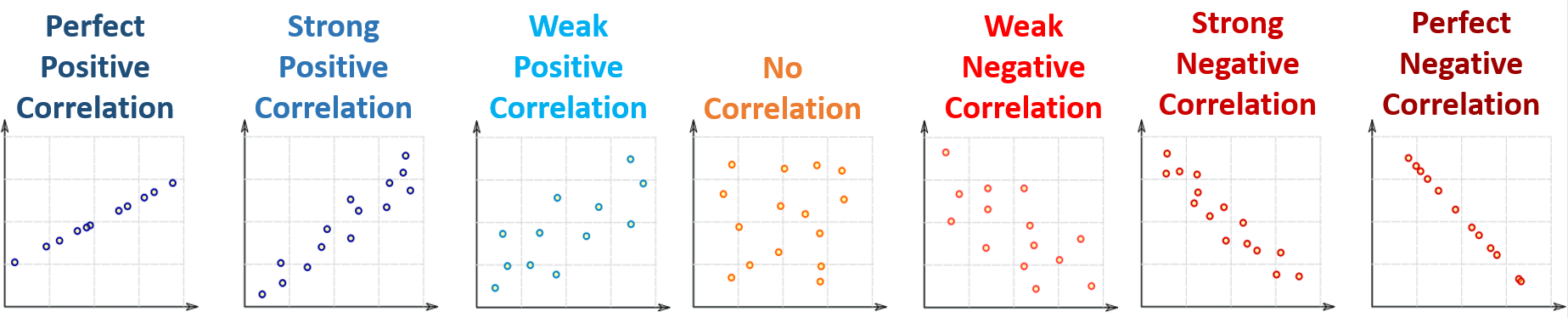
*Describing scatterplots*

Graphic inspired by *mathisfun.com*

Correlation: measures how two variables are **related**.

Positive correlations: as the x values increase, the y values also tend to **increase**.

Negative correlations: as the x values increase, the y values tend to **decrease**.



Graphic inspired by *mathisfun.com*

Least Squares Regression Line (LSRL): a straight line that roughly puts half of your data **above** it and half **below** it.

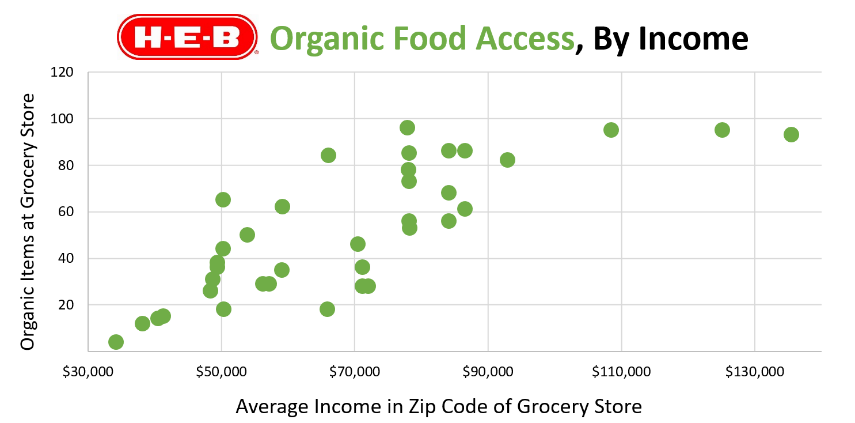
* More formal definition coming next lesson.

Strong correlations: data is **close** to the LSRL

* The LSRL is a **good model** for the data
* If you used the LSRL to predict new data, you would make **close predictions**.

Weak correlations: data is **far** from the LSRL

* The LSRL is a **‘meh’ model** for the data
* If you used the LSRL to predict new data, you would may be **quite a bit off**.

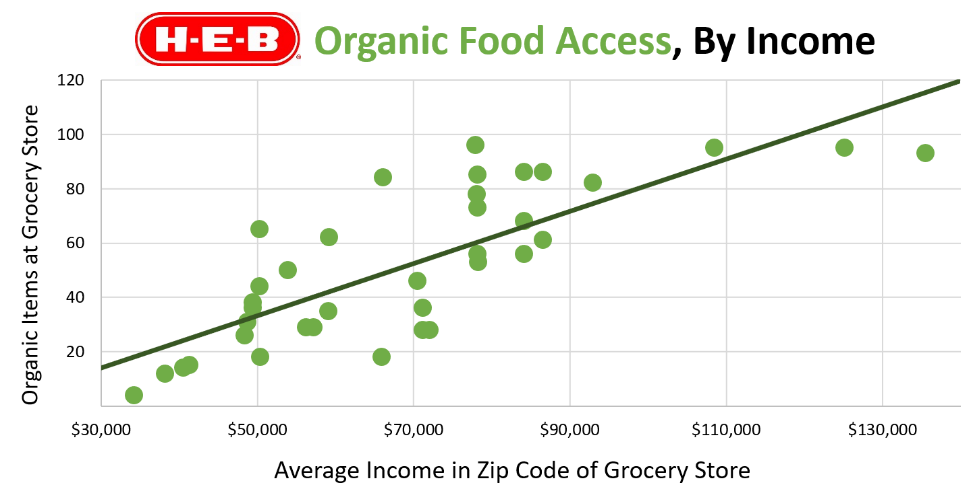
****

**Direction:**

**Positive**

**Strength:**

**Moderate**

****

**C – Context**

**D – Direction (positive/negative)**

**O – Outliers**

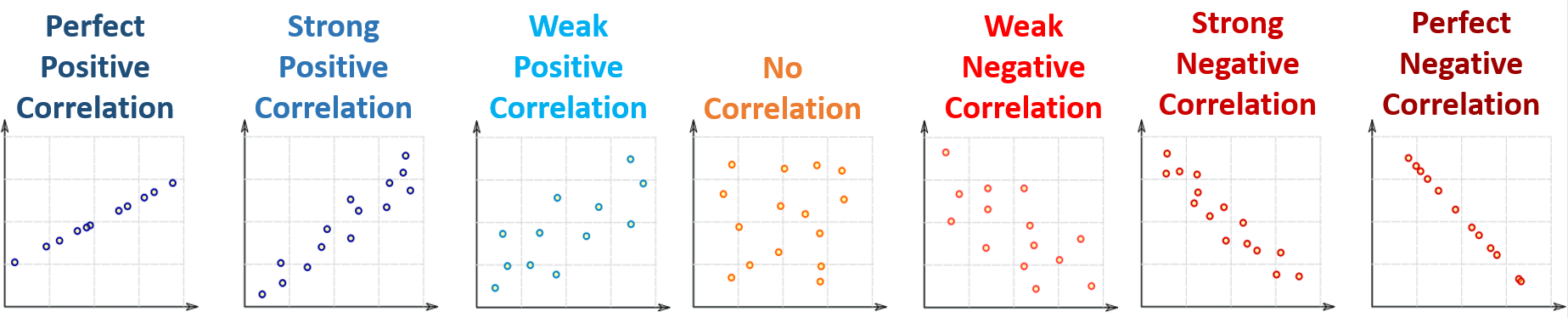
**F – Form (linear/non-linear)**

**S – Strength (strong/moderate/weak)**

*Put it all together: Describe the relationship between average household income in a zip code and the number of organic vegetables offered at the local grocery store…*

The relationship between average household income and the number of organic vegetables offered at local stores appears to be linear, positive, and moderately strong. There do not appear to be any outliers.

*Correlation coefficient (r)*



Graphic inspired by *mathisfun.com*

Correlation Coefficient (r): A number between **-1 and 1** that tells you the strength and direction of a correlation.

**Direction:**

Negative r value 🡪 **negative** correlation

Positive r value 🡪 **positive** correlation

**Strength:**

r close to 0 🡪 **weak** correlation

r close to -1, 1 🡪 **strong** correlation

**Lesson 3.1 Discussion**

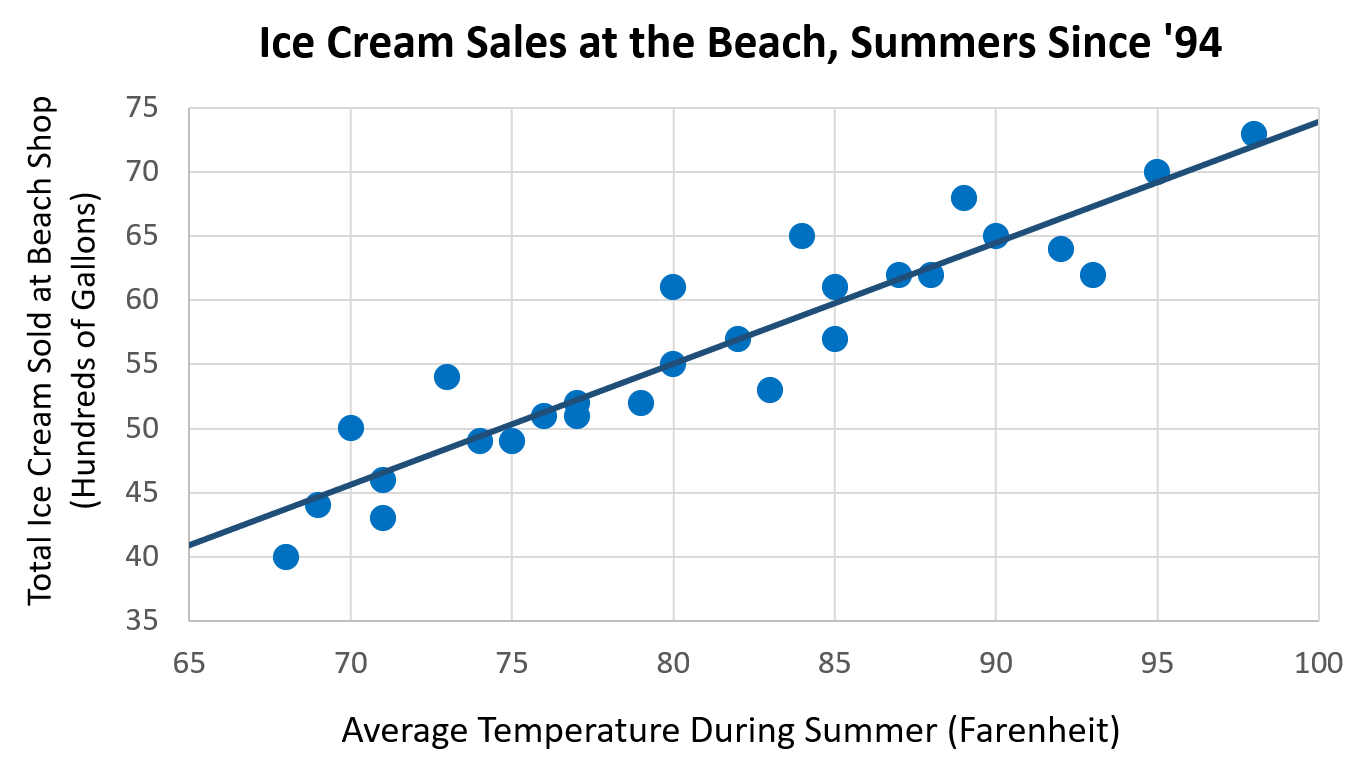
From 1994-2020, a statistician collected three pieces of data each summer at a beach:

1. The average temperature
2. The amount of ice cream sold at the beach shop

**Recommended discussion norms:**

[skewthescript.org/discussion-norms](https://skewthescript.org/discussion-norms)

1. The amount of drownings

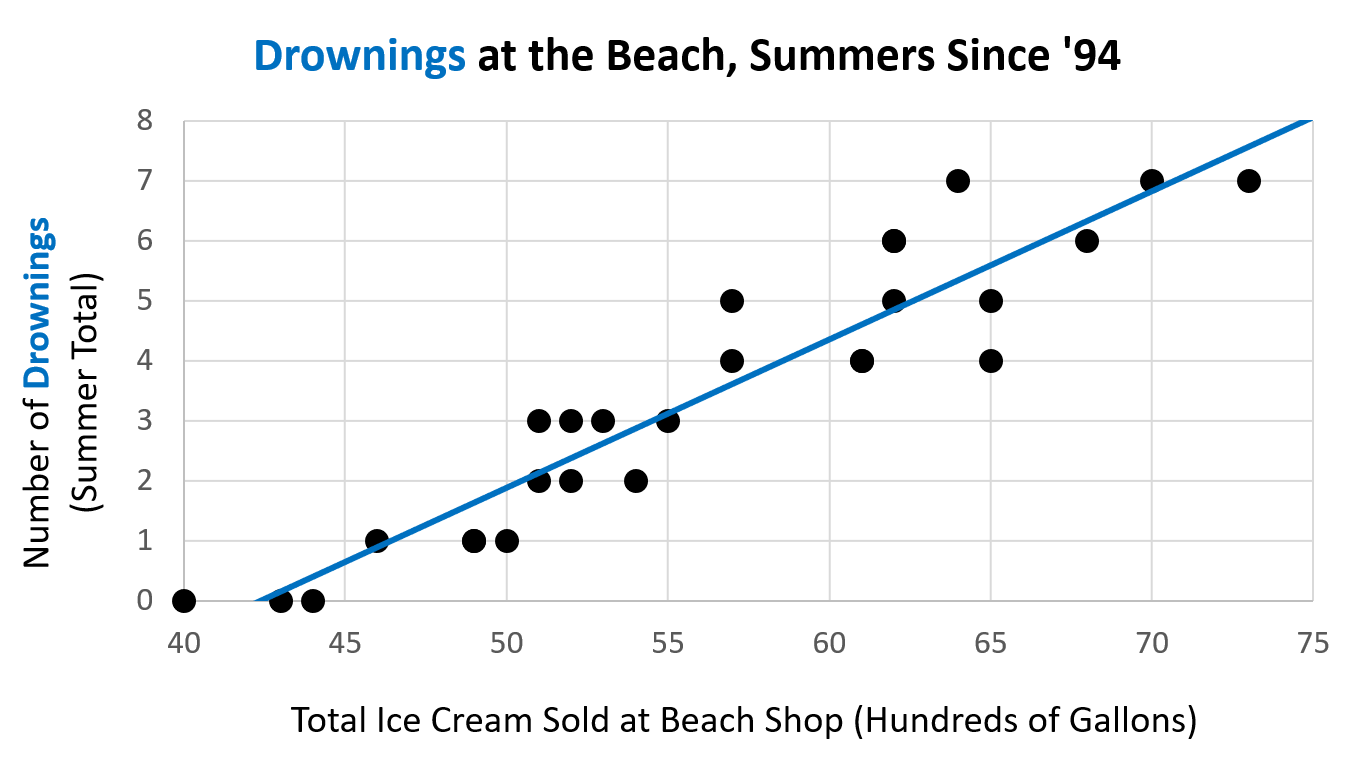


**A**

**A) Discussion Question:** Describe the relationship between temperature and ice cream sales. Does this relationship make sense? Why or why not?

The relationship between average temperature and ice cream sales appears to be linear, positive, and strong. There do not appear to be any outliers.

This relationship makes sense: People tend to crave a cool treat when temperatures are high. They won’t want ice cream as much when temperatures are low.

****

**B**

**B) Discussion Question:** Describe the relationship between ice cream sales and drownings. Does this relationship make sense? Why or why not?

The relationship between ice cream sales and drownings per summer appears to be linear, positive, and moderately strong. There do not appear to be any outliers.

**Teacher note:** Students will come up with all sorts of crazy answers here to explain this trend. You can discuss these explanations for a bit, but it’s important to navigate to the ultimate point:

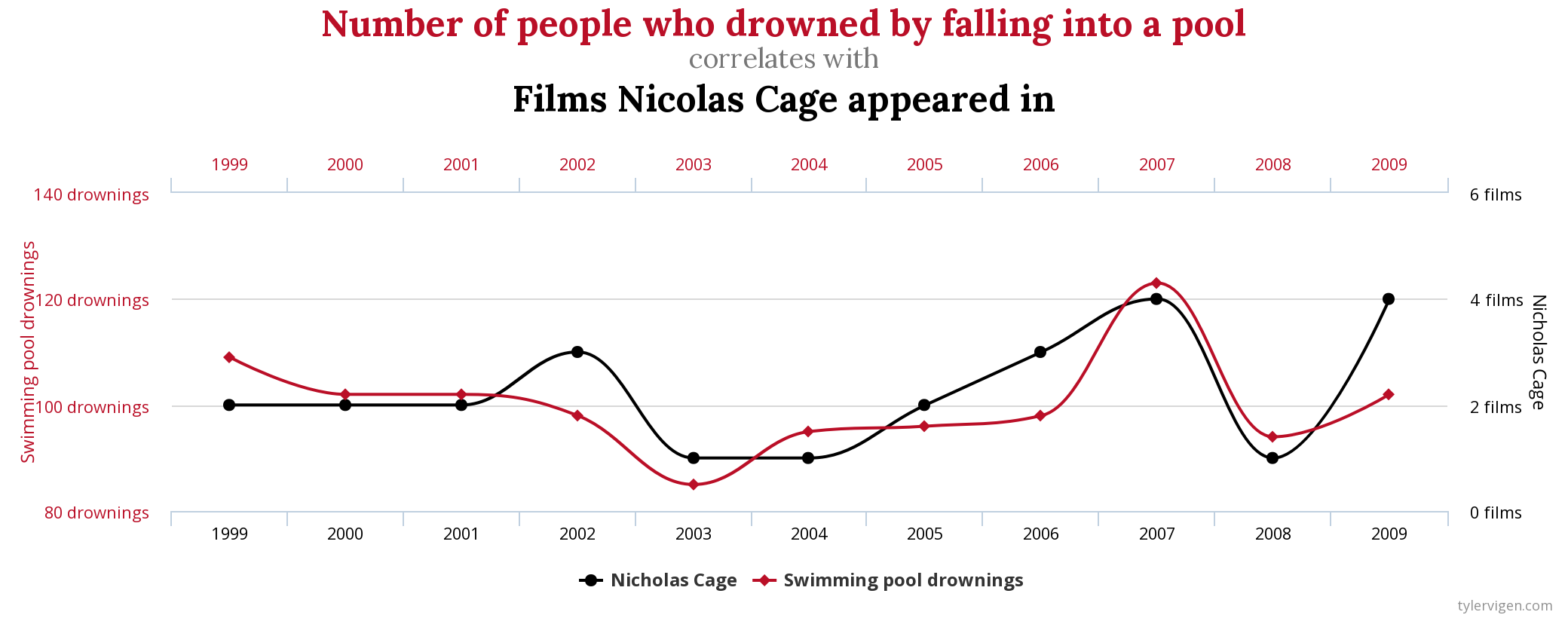
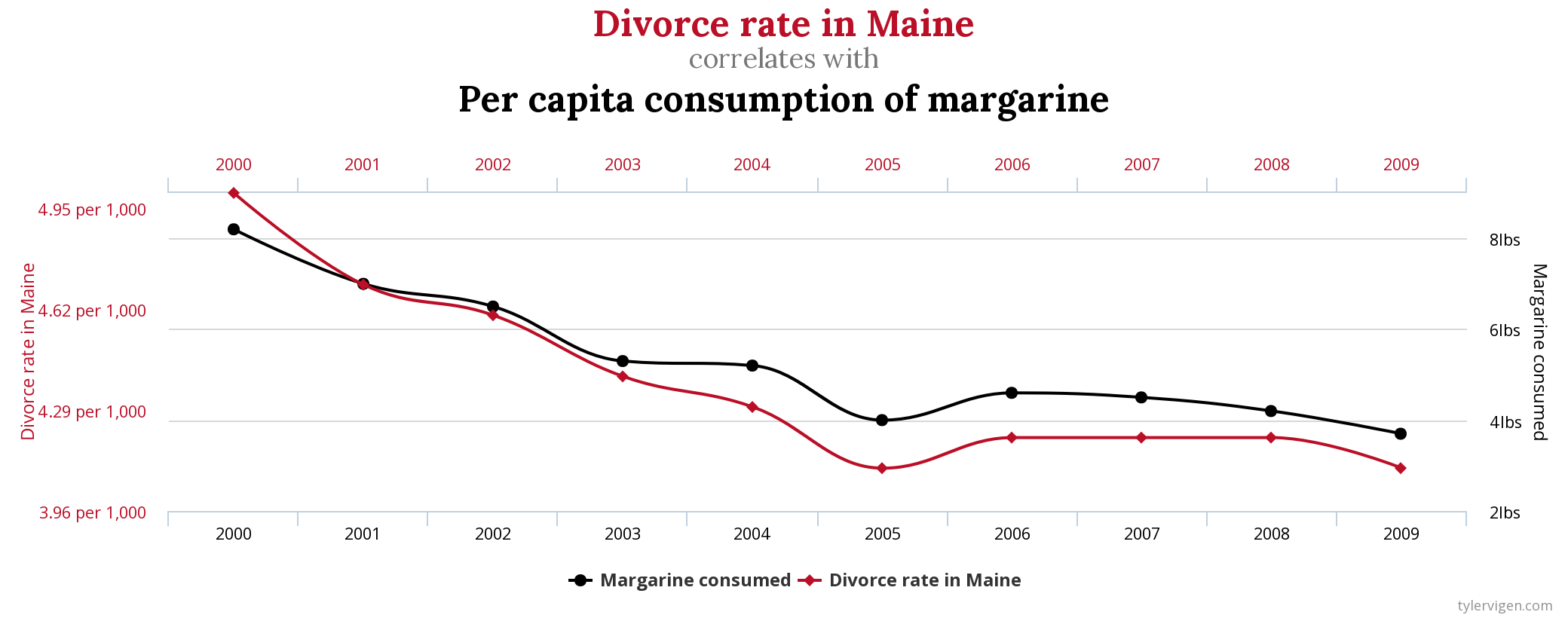
Ice cream sales ***do not cause*** drownings. The underlying variable of temperature is to blame here: as the summer gets hotter, more people buy ice cream. Simultaneously, more people go swimming and (unfortunately) drown. On colder days, fewer people buy ice cream and, simultaneously, fewer people swim and drown.

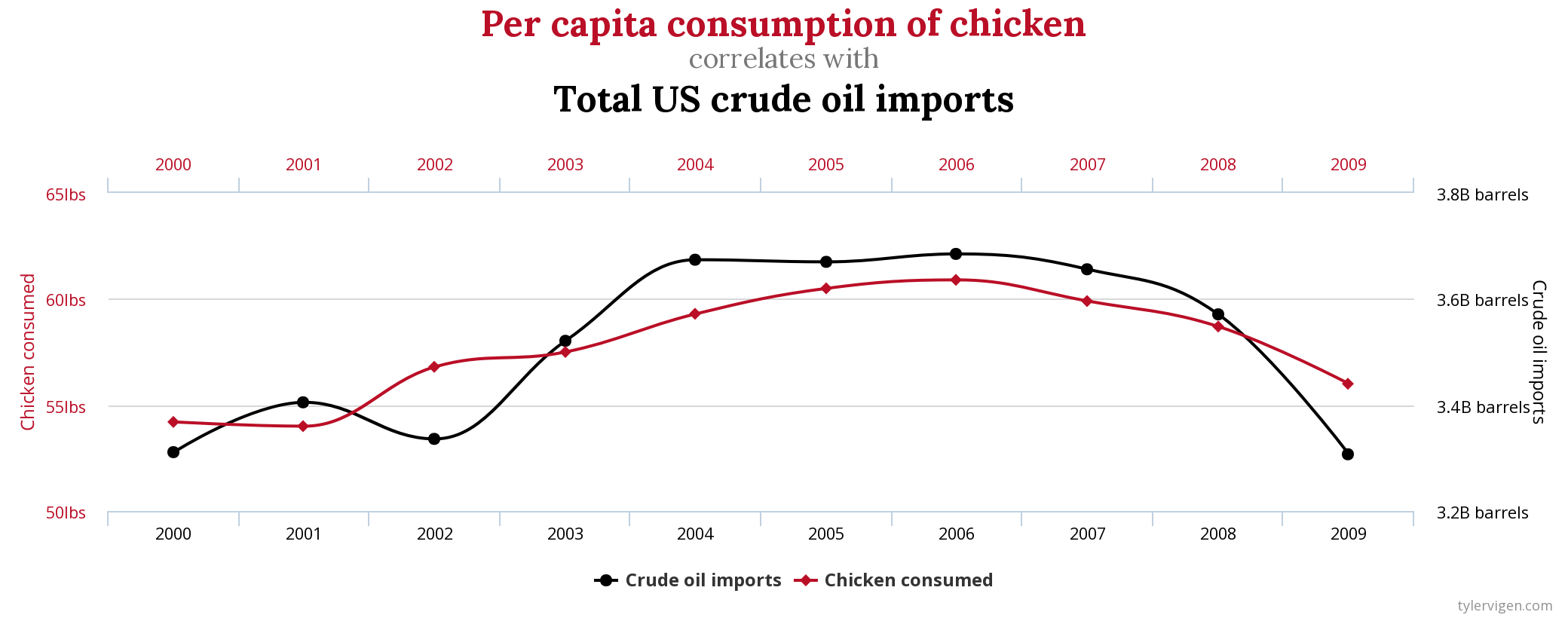
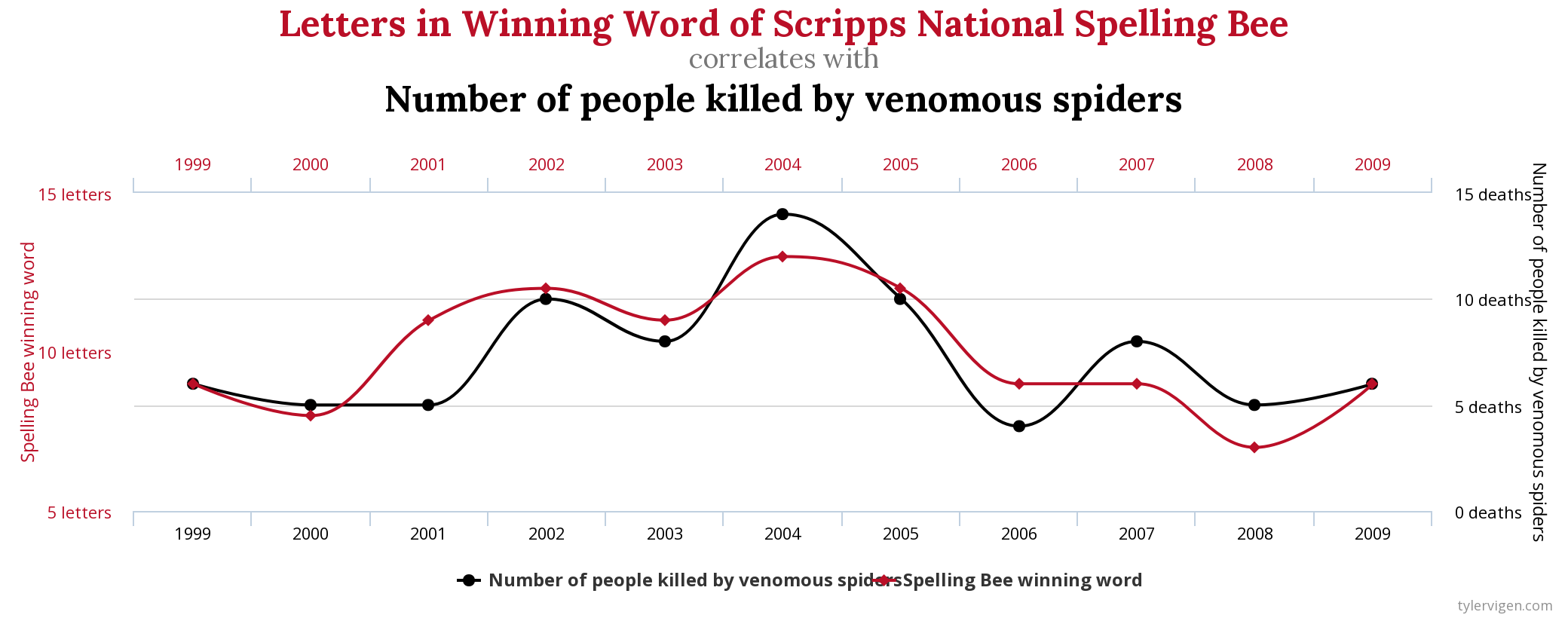
Example inspired by Ionica Smeets: <https://www.youtube.com/watch?v=8B271L3NtAw>

Big Idea:

**Correlation ≠ Causation**

*Coincidental Correlations, courtesy of Tyler Vigen:* [*https://www.tylervigen.com/spurious-correlations*](https://www.tylervigen.com/spurious-correlations)

****

****

**Lesson 3.1 Practice**

1. For some interactive practice, have students construct correlations and guess r-values using this applet: <http://digitalfirst.bfwpub.com/stats_applet/stats_applet_5_correg.html>. Source: *Digital First* project from Bedford, Freeman, & Worth publishers

2. In addition, we recommend providing practice exercises from your AP Stats textbook or from prior AP Stats exams. The following textbook sections and AP exam questions are aligned to the content covered in this lesson.

* [*The Practice of Statistics*](https://www.bfwpub.com/high-school/us/product/Practice-of-Statistics/p/1319113338) *(AP Edition),* 4th-6th editions: section 3.1
* [*Stats: Modeling the World*](https://www.pearson.com/store/p/stats-modeling-the-world/P100002559948/9780134685762) *(AP Edition),* 4th & 5th editions: chapter 6, 3rd edition: chapter 7
* [*Statistics: Learning from Data*](https://ngl.cengage.com/search/productOverview.do?N=201%204294918395&Ntk=NGL%7CP_EPI&Ntt=10570695941793977411692212681962783533&Ntx) *(AP Edition),* 2nd edition: section 4.1
* [AP Exam Free Response Questions (FRQs)](https://apcentral.collegeboard.org/courses/ap-statistics/exam/past-exam-questions): 2017 Q1 (part a), 2012 Q1, 2008 Q4 (parts a & b)

**Handout Key by statistics student Greyson Zuniga**