**Handout – Guided Notes**

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

**Explanatory and response variables**

Bivariate data: data with \_\_\_\_\_\_\_\_\_ variables.

Two quantitative variables are visualized in a \_\_\_\_\_\_\_\_\_\_\_\_\_\_

**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 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

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

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

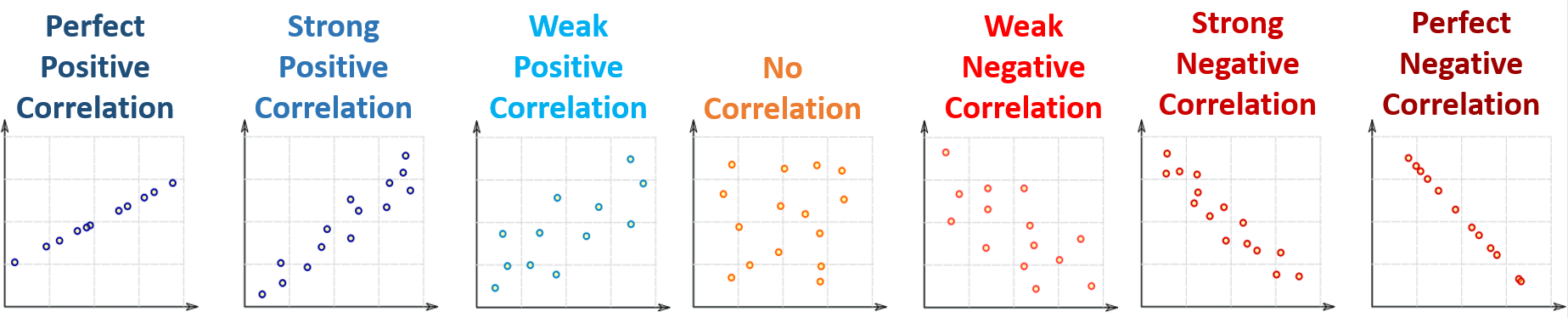
In this example, which variable is the explanatory variable? Why?

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

\*Dataset compiled by student Linda Saucedo, Fall 2019

**Describing scatterplots**

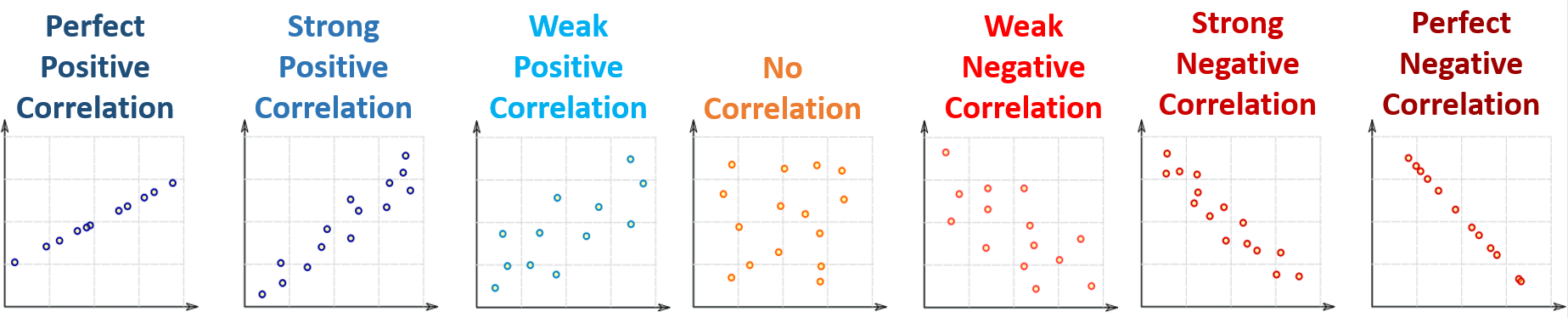
* **Direction**
* **Outliers**
* **Form**
* **Strength**
* **Context**



Correlation: measures how two variables are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

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

Negative correlations: as the x values increase, the y values tend to \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.



Least Squares Regression Line (LSRL):

a straight line that roughly puts half of your data \_\_\_\_\_\_\_\_\_\_\_\_ it and half \_\_\_\_\_\_\_\_\_\_\_\_\_\_ it.

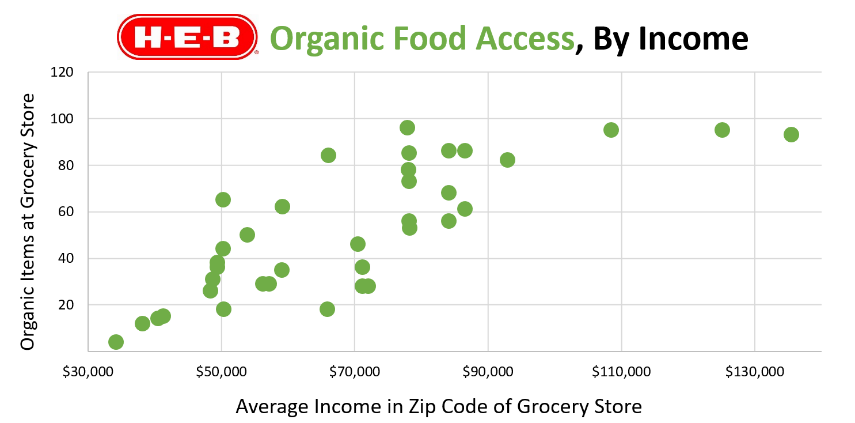
* More formal definition coming next lesson.

Strong correlations: data is \_\_\_\_\_\_\_\_\_\_ to the LSRL

* The LSRL is a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ for the data
* If you used the LSRL to predict new data, you would make \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

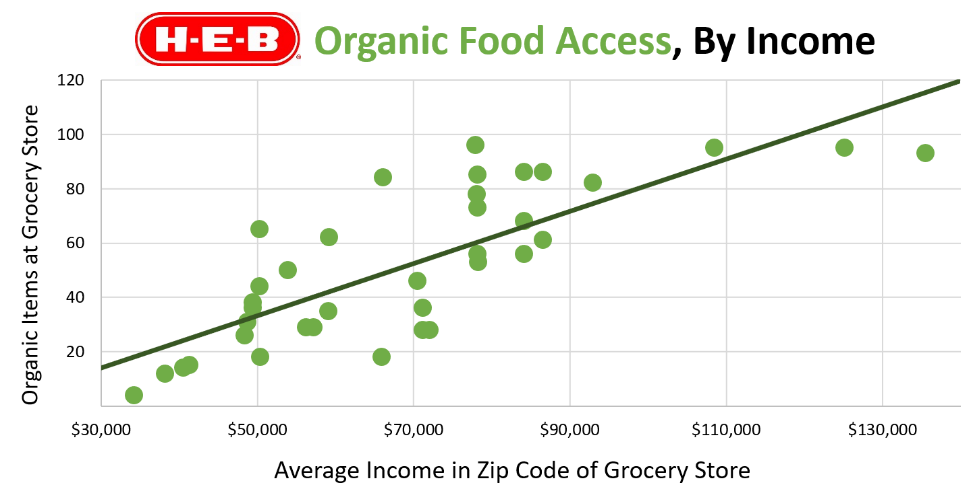
Weak correlations: data is \_\_\_\_\_\_ from the LSRL

* The LSRL is a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ for the data
* If you used the LSRL to predict new data, you would may be \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

****

**Direction:**

**Strength:**

****

**C – Context**

**D – Direction (positive/negative)**

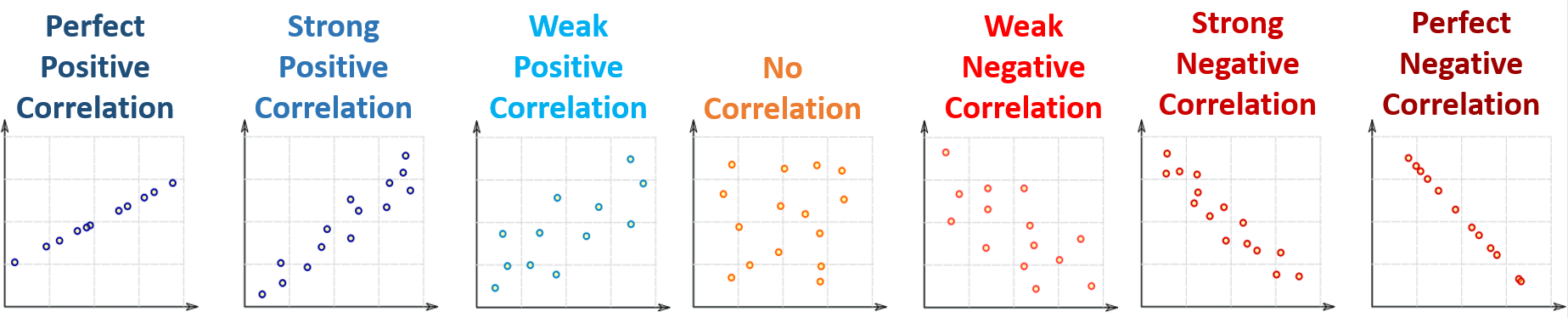
**O – Outliers**

**F – Form (linear/non-linear)**

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

**Put it all together:**

**Correlation coefficient (r)**



Graphic inspired by *mathisfun.com*

Correlation Coefficient (r):

A number between \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ that tells you the strength and direction of a correlation.

**Direction:**

Negative r value 🡪 \_\_\_\_\_\_\_\_\_\_\_\_\_\_ correlation

Positive r value 🡪 \_\_\_\_\_\_\_\_\_\_\_\_\_\_ correlation

**Strength:**

r close to 0 🡪 \_\_\_\_\_\_\_\_\_ correlation

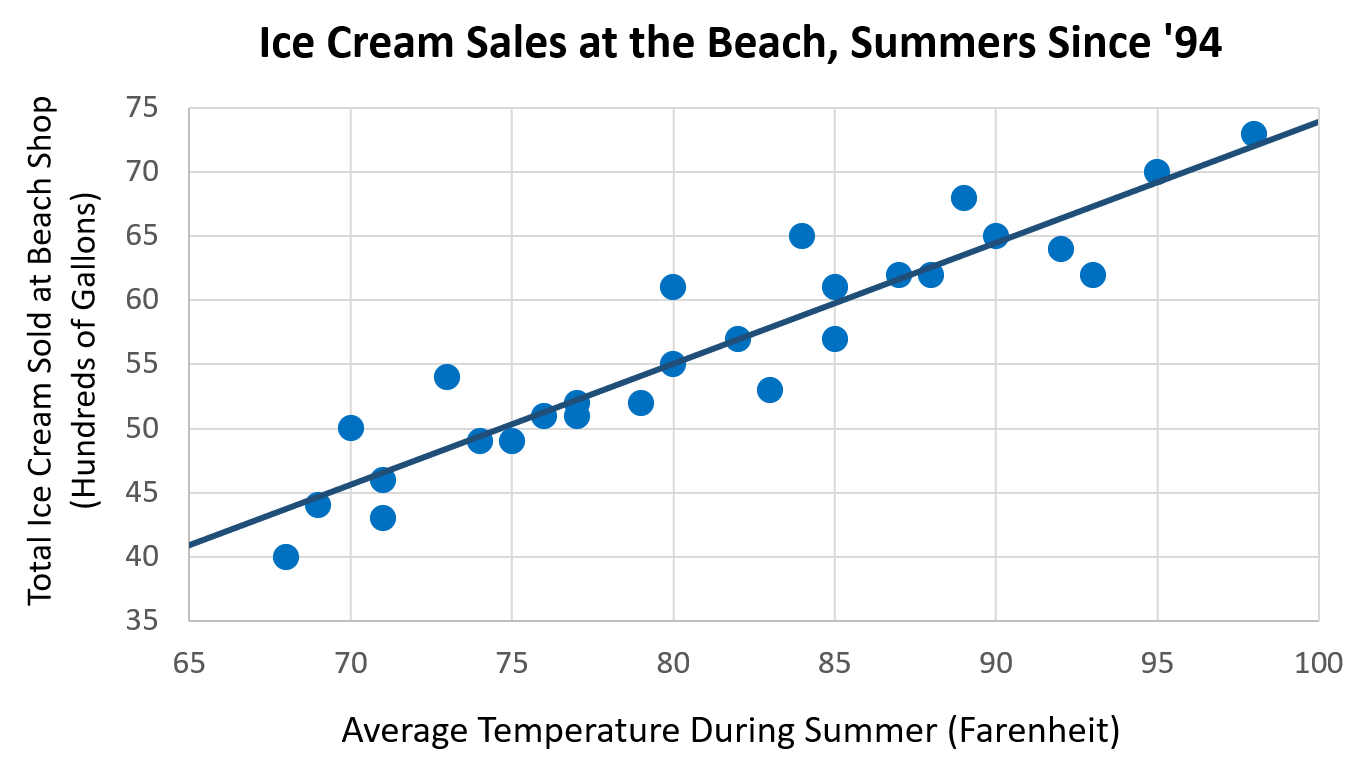
r close to -1, 1 🡪 \_\_\_\_\_\_\_\_\_\_ correlation

**Discussion**

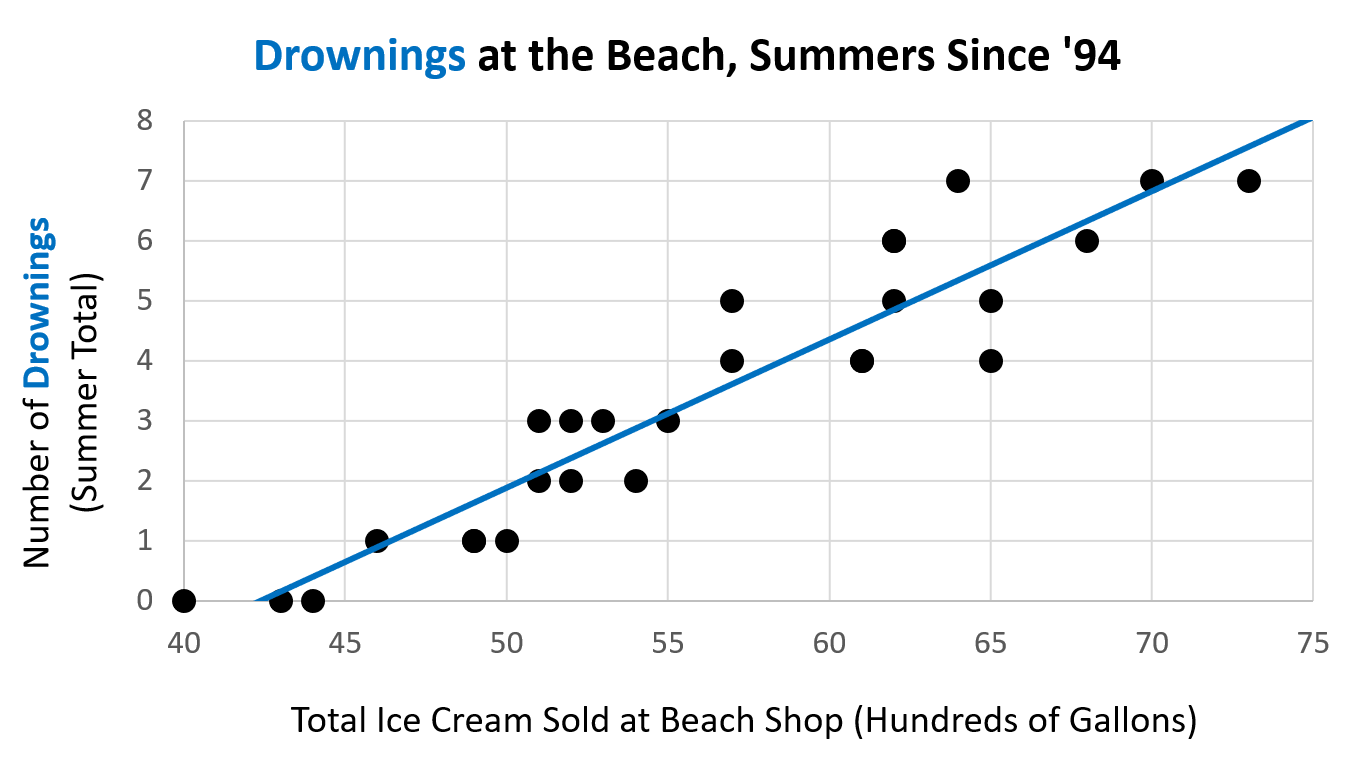
From 1994-2020, a statistician collected three pieces of data each summer at a beach:

1. The average temperature
2. The amount of ice cream sold at the beach shop
3. 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?

****

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

**B**

Big Idea: **Correlation \_\_\_ Causation**