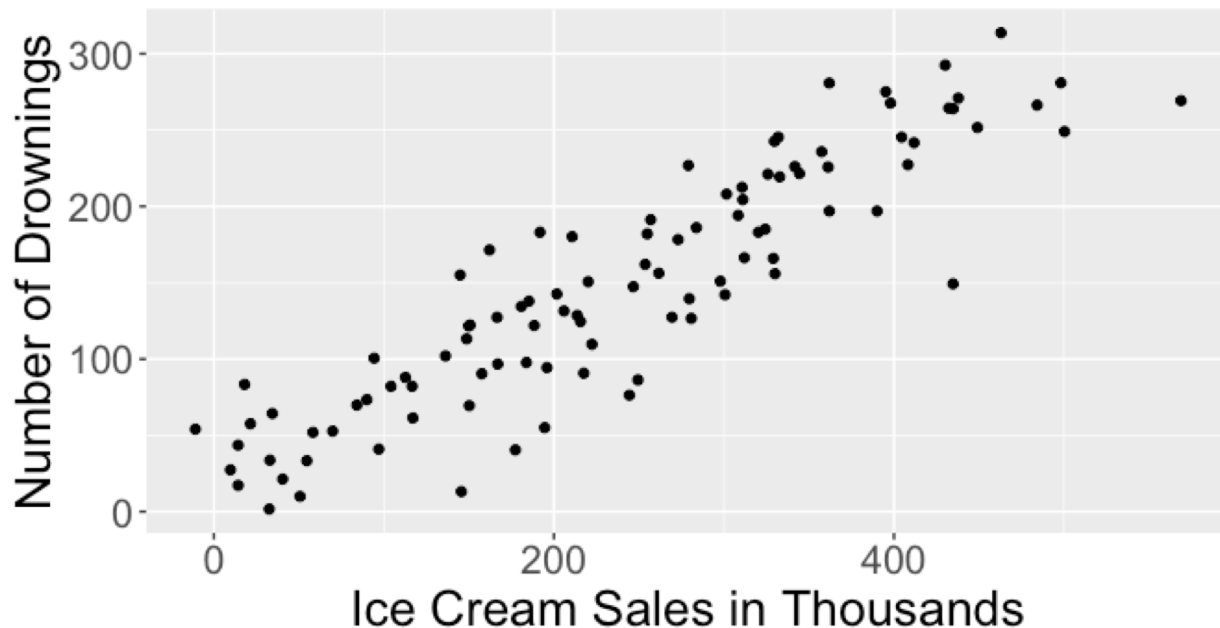# Confounding

Example: Ice Cream Sales

Suppose we wish to examine the relationship between number of drownings and ice cream sales. (Each point represents a different day).

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  22.02858    6.99707   3.148  0.00218 **
icecreamsales  0.52688    0.02543  20.717  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- The output even suggests that ice cream sales is significantly and positively associated with number of drownings (Beta=0.5269, p-value<0.001). In particular, the model estimates that as ice cream sales increases by 1 unit (in this case, 1 thousand dollars), the number of drownings is expected to increase by 0.53 on average.

- Why can't we say that an increase in ice cream sales cause more drownings?

Example: Ice Cream Sales (cont.)

If we include temperature in the plot, what can we see happening?
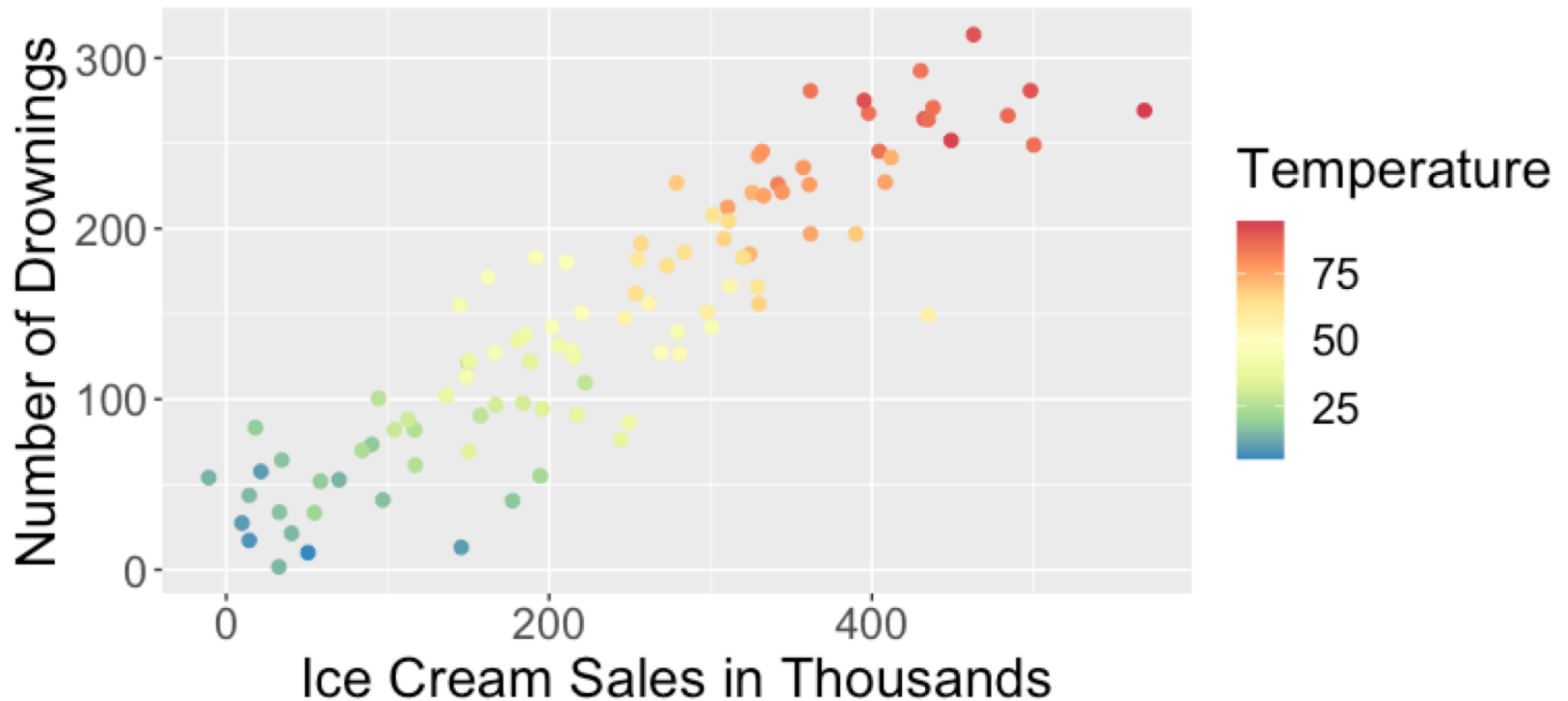
- Logically it makes sense that temperature is the driving force here. As temperature increases, it is likely makes people more interested in purchasing ice cream. It also makes people more likely to swim, which would result in a higher number of swimming-related incidents.

- A confounding variable is one that is linked to both the explanatory variable and the response variable.

Example: Ice Cream Sales (cont.)

- If we control for temperature, are ice cream sales still significantly associated with number of drownings?

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.30286    4.26849  -0.071    0.944
icecreamsales -0.03627    0.04176  -0.868    0.387
temperature    3.15915    0.21981  14.372   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- After controlling for temperature, ice cream sales are no longer significantly associated with number of drownings (Beta=-0.036, p-value=0.387).

Confounding

- A potential confounding variable is one that we expect to be related to both the explanatory and response variable.

- If our relationship of interest remains unchanged after including potential confounding variables, it can make your overall story more powerful. While you won't be able to establish causation - it does get you closer to understanding how your variables relate to one another.

Confounder?

- Suppose that we find that there is a significant association between number of fire fighters responding to a fire and the total damage done in a fire. What is a potential confounding variable in this relationship? (From the video!)

- Suppose we find that there is a significant association between activity level and energy. What is a potential confounding variable in this relationship?

- Suppose we want to understand the relationship between size of home and blood pressure. What is a potential confounding variable in this relationship?