Lab 2F - The Titanic Shuffle

Directions: Follow along with the slides and answer the questions in **bold** font in your journal.

## Previously ...

* In the previous lab, we learned that by using a do-loop and the shuffle function, we could simulate randomly shuffling our data many times.
* This helps us determine how likely it is that a difference between groups is due to chance.
* For this lab, will extend these ideas to *numerical* variables by using random shuffling and numerical summaries.
* The question we will investigate in this lab is:

*Is there any evidence to suggest that wealthier passengers on the Titanic were more likely to survive than poorer passengers.*

* We will consider wealthier passengers to be those that paid a higher fare for their ticket.

## The Titanic

* The Titanic was a ship that sank en route to the U.S.A. from England after hitting an Iceberg in 1912.
  + At the time, it was claimed that the Titanic was *unsinkable* ... it wasn't ... because it did.
* Use the data function to load the titanic passenger and survival data.
* Create a boxplot of the fares paid by passengers and facet the plot based on whether the passenger survived or not.
  + **Based on the plot, do you believe richer passengers were more likely to survive? Explain why and describe how certain you are of being correct.**

## The search begins!

* Start your analysis by calculating how much more the *typical* survivor paid than the *typical* non-survivor in our data.
  + Based on the distributions of fares paid, which numerical summary that describes the *typical* value might be preferred?
* **What was the *typical* fare paid by survivors? Non-survivors? How much more did the typical survivor pay?**

## Do the shuffle!

* Use the do and the shuffle functions to shuffle the passenger's survival status 500 times.
  + Use the previous lab if you need some help on how to do this.
  + For each shuffle, compute each group's median fare paid.
  + Assign your shuffled data the name shuffles.
* After shuffling your data, use the mutate function to create a variable called diff to the shuffled medians you just calculated. (Assign your mutated data the name shuffles again).

## Put your simulations to use

* **By using your shuffled data, answer the research question we posed at the beginning of the lab.**

*Is there any evidence to suggest that wealthier passengers on the Titanic were more likely to survive than poorer passengers.*

* **Write up your answer as a statistical analysis. Create a plot and explain how the plot supports your conclusion. Be sure to also explain why shuffling your data is important.**