Case Study 1

In order to study gender bias in promotion decisions, researchers took 48 resumes and randomly assigned them to 48 bank managers who were asked whether they would promote this candidate or not. The resumes were identical except the name: 24 of them had names generally associated with women, 24 of them had names generally associated with men.

The data from this study can be found in the `promote` data frame in the `stat20data` library. Did gender played a role in promotion decisions?

1. Write down an expression for the null and alternative hypothesis. Is H_0 an "independence" null hypothesis or a point null hypothesis?

2. Write an `infer` pipeline to compute the observed test statistic.

3. What type of plot can be used to visualize the observed data? Sketch a possibility for what the observed data might look like if it were *consistent* with the null hypothesis.

4. Write an `infer` pipeline to construct and save the null distribution of statistics.

5. Compute the p-value using an `infer` pipeline with the null distribution from the previous part.

6. Interpret the p-value. What does it say about the consistency between the null hypothesis and the observed data?

Case Study 2

In the 2016 NBA season, it was noted that professional basketball player Steph Curry had a remarkable basket-shooting performance beyond 30 feet. The `curry` data frame within the `stat20data` library contains his long range shooting performance across 27 attempts. By comparison, the long range shooting percentage of NBA players that season was 7.5%.

Assess whether this data is consistent with the notion that Steph Curry has a long range shooting percentage that is no different from the average NBA player. Said another way, assess just how remarkable Curry's shooting performance was in 2016.

7. Write down an expression for the null and alternative hypothesis. Is H_0 an "independence" null hypothesis or a point null hypothesis?

8. Write an `infer` pipeline to compute the observed test statistic.

9. Write an `infer` and `ggplot2` pipeline to construct a plot featuring 9 subplots, each one featuring a visualization of a data set generated under the null hypothesis.

10. Does your visualization of the observed data from the previous part look like it could be one of these plots?

11. Visualize the null distribution using an `infer` pipeline.

12. Sketch the null distribution that you found below, labeling axes and shading the area corresponding to areas considered "extreme".

13. What does your visualization say about the consistency between the null hypothesis and the observed data?