1. **Australian Open:** The Australian Open is an annual tennis tournament in Melbourne, Australia. 128 players take part in the singles tournament and are matched up in the first round based on their rank at the time. We have [data](https://www.kaggle.com/datasets/valentinfiorenza/tennis-data) from the men’s tournament in the years 2008-2022, with variables for both the winner and loser of each match, along with information such as what hand they play with and their Association of Tennis Professionals (ATP) rank.

**Goal:** To determine if there is a relationship between the difference in rank and whether the higher rank won in the whole tournament

## higher\_rank\_won

## rank\_diff\_cat No Yes

## Low Difference 230 429

## Medium Difference 141 532

## High Difference 115 452

Table 1: 2x3 comparison of rank difference to whether the higher rank won over the whole tournament

* 1. Using Table 1, what is the number of times that there was a "High Difference" in the players' ranks?
  2. Calculate the expected values for each rank\_diff\_cat and higher\_rank\_won pair.
  3. Using the expected values you have created, calculate the chi-squared value:
  4. Determine if there is a relationship between the level of rank difference and whether the higher rank won at:
     1. 1% level of significance
     2. 5% level of significance

**Goal:** Now that we have worked through an example as a group, repeat the same steps to determine if there is a relationship between the difference in rank and whether the higher rank won in the whole tournament

## higher\_rank\_won

## rank\_diff\_cat No Yes

## Low Difference 25 65

## Medium Difference 2 10

## High Difference 1 2

Table 2: 2x3 comparison of rank difference to whether the higher rank won in the Quarterfinals, Semifinals, and Finals

1. Using Table 1, what is the number of times that there was a "High Difference" in the players' ranks?
2. Calculate the expected values for each rank\_diff\_cat and higher\_rank\_won pair.
3. Using the expected values you have created, calculate the chi-squared value:
4. Determine if there is a relationship between the level of rank difference and whether the higher rank won at:
   * 1. 1% level of significance
     2. 5% level of significance