Data analysis plan post-study completion:

1. Files titled: prep_study1
2. R code: prep_study1_code
3. Variables (with name in code): age, gender, race/ethnicity (race & race_other), MTurkID (turkid), task comprehension (wrong_ task_comp), condition (cond), number of incorrect for payment schemes (count_pay_comp), choice of payment scheme (pay_choice), choice to prepare (prep_choice), time spent preparing (prep_time), score, perceived percentile rank (conf), perceptions of better performing gender (percep_score), perceptions of gender that spent more time preparing (percep_prep), perceptions of differences in tournament entry (percep_entry), risk, income, highest level of education completed (edu)
4. Send out bonus payments
    a. Randomize each participant to another person (who chose competition)
        i. Tournament
            1. Randomly sort data frame
            2. Split in half (create two new data frames)
            3. Subtract 1 person's score from another's in respective round
            4. Create new column with payoff where negative values are recoded to 0, 0 values are recoded to… else (e.g. pos scores) multiply by round score
            5. Create new column: If they win – copy and paste bonus from bonus column, else bonus = 0
            6. Sort by bonus col (with 0 at bottom)
        ii. Confidence measures
            1. If question 1 (percentile rank compared to all others) is chosen randomly:
                a. One option: https://blog.exploratory.io/how-to-calculate-percentile-ranks-in-r-and-exploratory-85ca95856ceb
                b. Another option: https://stackoverflow.com/questions/25760504/use-dplyrpercent-rank-to-compute-percentile-ranks-within-group
            2. If question 2 (Do you think men or women correctly solved more multiplication problems on average?) is chosen randomly, calculate mean for men and women. Create new column with correct (=1) or not (=0). Multiply score by correct or not.
            3. If question 3 (Do you think men or women spent more time practicing/studying before completing the multiplication task?) is chosen randomly, '' above
            4. If question 4 (Do you think men or women chose the tournament payment option more often?) is chosen randomly, '' above
    b. Once bonuses are calculated, go to study on dashboard and click "Grant bonuses," then under there – click "try our new feature of granting workers bonus amounts"

Commented [RK1]: Another option is to write lines of code that subtract col n-1 from col n: https://www.reddit.com/r/rstats/comments/3pgjgj/how_to_create_a_new_column_whose_value_is_the/
****

Commented [RK2]: Ignore the group by part – unnecessary

a. Please note that your CSV file should have Worker IDs in the first column and the total amount you want to bonus in the second column. Your file should not have headers
b. Message to workers: "Thank you for participating in our study on decision-making. You have been awarded $X in bonuses."
c. If you already gave a Worker a bonus and you want to give this Worker another bonus you need to enter the total bonus amount.

5. Clean data
    a. Recoding scheme
        i. Gender
            1. Man = 0
            2. Woman = 1
        ii. Competition choice
            1. Tournament = 0
            2. Piece-rate = 1
        iii. Prep choice
            1. No = 0
            2. Yes = 1
        iv. Condition
            1. Control = 0
            2. Preparation = 1
        v. Perceptions of better performing gender
            1. Men = 0
            2. Women = 1
        vi. Perceptions of gender that spent more time preparing
            1. Men = 0
            2. Women = 1
        vii. Perceptions of differences in tournament entry
            1. Men = 0
            2. Women = 1
        viii. Race/ethnicity
            1. White = 0
            2. Black/African American = 1
            3. Hispanic/Latino = 2
            4. East Asian = 3
            5. Middle Eastern = 4
            6. Native American = 5
            7. Pacific Islander = 6
            8. Mixed = 7
            9. Other = 8
        ix. What is your highest level of education completed?
            1. Less than a high school degree = 0

Commented [RK3]: potentially include task comp if dichotomous var

2. High school diploma = 1
3. Vocational training = 2
4. Some college = 3
5. Bachelor's degree = 4
6. Graduate degree = 5

x. Ownincome
1. Less than $10,000 = 0
2. $10K to 20K = 1
3. $20K – 30K = 2
4. …
5. More than $200K = 11

6. Exploring data:
   a. Plot graphs of the following:
      i. Proportion of people that chose to compete
      ii. Proportion of women and men
         1. Then proportion of men and women that competed
         2. Then proportion of men and women that competed conditional on condition
      iii. Count of people that chose to prepare
         1. Then proportion of men and women that prepared
         2. Then proportion of men and women that competed conditional on preparation choice
      iv. Proportion of race/ethnicity (by gender)
      v. Proportions by income (by gender)
      vi. Level of education completed (by gender)
      vii. Confidence measure (by gender)
      viii. Proportions for 3 perceptions questions
      ix. Risk measure (by gender)
      x. Scatterplot of time spent preparing (by gender)
      xi. Scatterplot of age (by gender)
      xii. Proportion of pay comprehension (by gender)
      xiii. Score (by gender)