# Tasks

*Tip:* I recommend in the code block commenting which number the code answers. It will help me find your answers more readily. Something as simple as # 3B helps me.

1. Formulate your Repo
   1. Tori: The client wants this to be reproducible and know exactly what you did. There needs to be an informative Readme, complete with several sections, as referenced in Live Session. Give contact information, session Info, and the objective of the repo at least.
   2. An: You have a large data set, and it needs its own Codebook, formatted in an approachable way. Make sure you describe peculiarities of the data by variable and what needs transforming. However, do not make it too long either.
   3. Rajat: Create a file structure that is accessible and transparent. Document it in the root directory, ideally in the Readme.
2. Clean your Raw Data
   1. Andy: Read the csv into R and take a look at the data set. Output how many rows and columns the data.frame is.
   2. Andy: The column names are either too much or not enough. Change the column names so that they do not have spaces, underscores, slashes, and the like. All column names should be under 12 characters. Make sure you’re updating your codebook with information on the tidied data set as well.
   3. Andy: Some columns are, due to Qualtrics, malfunctioning.
   4. Andy: Make sure your columns are the proper data types (i.e., numeric, character, etc.). If they are incorrect, convert them.
3. Preliminary Analysis
   1. An: Remove all observations where the participant is under age 18. No further analysis of underage individuals is permitted by your client. Remove any other age outliers as you see fit, but be sure to tell what you’re doing and why.
   2. An: Please provide (in table format or similar), descriptive statistics on at least 7 variables (age, Income, etc.). Create a simple histogram for *two* of them. Comment on the shape of the distribution in your markdown.
   3. Tori: Give the frequencies (in table format or similar) for Gender, Education, and Occupation. They can be separate tables, if that’s your choice.
   4. Tori: Give the counts (again, table) of management positions.
4. Deeper Analysis and Visualization

*Note:* You should make all of these appealing looking. Remember to include things like a clean, informative title, axis labels that are in plain English, and readable axis values that do not overlap.

* 1. An: Create bar charts in ggplot or similar. The bars should be in descending order, Use any color palette of your choice other than the default.
  2. An: Is there a relationship between Age and Income? Create a scatterplot and make an assessment of whether there is a relationship. Color each point based on the Gender of the participant. You’re welcome to use lm() or similar functions to back up your claims.
  3. Jodi: What about Life Satisfaction? Create another scatterplot. Is there a discernible relationship there to what?

1. Presentation
   1. Jodi: Create framework for presentation
   2. Slide assignments (to be assigned next meeting):
      1. Title/Intro (*name*)
         1. Authors all listed
      2. Presentation Outline ()
         1. Business Objectives
         2. Data Sourced
         3. Methodology
         4. Evaluation/Results
         5. Summary
      3. Business Objectives ()
      4. Data Source ()
         1. Where you got it
         2. Basic Statistics (EDA)
      5. Methodology ()
         1. Steps
         2. Workflow
      6. Evaluation/Results ()
         1. Tell me the percentages and why
         2. Show me graphs with explanations
         3. The top three factors that contribute to turnover
         4. Tell me about any job role specific trends that may exist in the data set
         5. Provide any other interesting trends and observations from your analysis
         6. Other things to consider?
      7. Summary ()
         1. Insights
         2. Recommendations
         3. Improvements
      8. Questions ()

**Variables to EDA by person:**

* Jodi
  + EducationField
  + Education
  + TrainingTimesLastYear
  + TotalWorkingYears
  + YearsAtCompany
  + YearsInCurrentRole
  + YearsSinceLastPromotion
  + YearsWithCurrManager (edited)
* Tori
  + Gender
  + Age
  + DistanceFromHome
  + EnvironmentSatisfaction
  + JobSatisfaction
  + MaritalStatus
  + RelationshipSatisfaction
  + WorkLifeBalance
* An
  + OverTime
  + DailyRate
  + HourlyRate
  + MonthlyIncome
  + MonthlyRate
  + PercentSalaryHike
  + StockOptionLevel
* Andy
  + Department
  + JobRole
  + JobInvolvement
  + JobLevel
  + NumCompaniesWorked
  + PerformanceRating
  + BusinessTravel