1. The data `Dem_data.csv` contains 5 variables:

   - `district`: the congressional district (1 to 435)
   - `Year`: the year of voting (t_{ij} below),
   - `Percent_Dem`: the percentage of voters that voted Democrat (Y_{ij} below)
   - `Percent_Dem_prev`: the percentage of voters that voted Democrat in the previous election, (X_{ij1} below)
   - `Incumbency`: 1 if the incumbent was a Democrat, 0 for open seats, and -1 for a Republican incumbent (X_{ij2} below).

   The model we model we want to run is:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 X_{ij1} + \beta_3 X_{ij2} + b_{0i} + \varepsilon_{ij} \tag{1}$$

   where the variable are defined above. You should check if any of the variables need to be included as class variables (or some other non-linear transformation like X^2 or $\log X$).

   Missing values correspond to congressional races that were not contested. The dataset contains years 1984, 1986, 1988, and 1990 where, for example, for 1984 the `Percent_Dem_prev` variable represents that percentage of democrat voters in the 1982 election. District lines are redrawn every ten years and change for years ending in '2' (e.g., 1980 to 1982 would have different district lines and cannot be compared).

   (a) **Qualitatively** assess whether the missing data mechanism is MCAR, MAR or MNAR. Be as specific as possible using the vocabulary we discussed in class.

   (b) Regardless of your assessment in (a), **quantitatively** test MCAR vs MAR.

   (c) Construct two "bad" imputation procedures and one "good" imputation procedure for the uncontested elections.

   (d) Fit a model to each of the imputation methods in (c).