**median\_wages.csv Description:**

This data set contains outcome median wage measures at the firm-year level. Note that if the firm has zero total employees in a given year, that firm-year observation will not appear in median\_wages.csv.

In particular, the variables in the data set are:

* payer\_tin\_w2\_max:
  + The firm id of the observation
* tax\_yr:
  + The tax year of the observation
* emps
  + The number of employees at the firm in the firm-year grouping
* med\_wages
  + The median wages of all employees (i.e., pooled) in the firm-year grouping. If there are fewer than 3 workers in this grouping, median wages are set to missing.
* males
  + The number of male workers at the firm in the firm-year grouping
* med\_male\_wages
  + The median wages of male employees in the firm-year grouping. If there are fewer than 3 males in this grouping, median male wages are set to missing.
* females
  + The number of female workers at the firm in the firm-year grouping
* med\_female\_wages
  + The median wages of female employees in the firm-year grouping. If there are fewer than 3 females in this grouping, median female wages are set to missing.
* inventors
  + The number of inventors at the firm in the firm-year grouping
* med\_inventor\_wages
  + The median wages of inventors in the firm-year grouping. If there are fewer than 3 inventors in this grouping, median inventor wages are set to missing.
* noninventors
  + The number of non-inventors at the firm in the firm-year grouping
* med\_noninventor\_wages
  + The median wages of non-inventors in the firm-year grouping. If there are fewer than 3 non-inventors in this grouping, median non-inventor wages are set to missing.
* stayers
  + The number of firm stayers at the firm in the firm-year grouping
* med\_stayer\_wages
  + The **\*\*CONTEMPORANEOUS\*\*** median wages of stayers in the firm-year grouping. If there are fewer than 3 stayers in this grouping, median stayer wages are set to missing.
* medindiv\_stayer\_wages
  + The median wages of the stayer in the associated year from the given firm-year grouping who was the median worker in the application year. If there are multiple individuals (due to a tie in median wages, or even number of workers) then this variable contains the mean of the stayers’ wages in the associated year. If there are fewer than 3 stayers in this grouping, median stayer wages are set to missing.
    - For example, say there are 5 stayers at firm 10 in 2009. Firm 10 applied for a patent in 2005, and at that time there were 10 workers. We restrict ourselves to the 5 workers who are present in 2009, find the median individual among these 5 workers according to their 2005 wages, and then return the 2009 wages for this median individual as the “medindiv\_stayer\_wages” variable.
* medindiv\_stayer\_appyr\_wages
  + The median wages of the stayer in the application year from the given firm-year grouping who was the median worker in the application year. If there are multiple individuals (due to a tie in median wages, or even number of workers) then this variable contains the mean of the stayers’ wages in the application year. If there are fewer than 3 stayers in this grouping, median stayer application year wages are set to missing.
    - For example, say there are 5 stayers at firm 10 in 2009. Firm 10 applied for a patent in 2005, and at that time there were 10 workers. We restrict ourselves to the 5 workers who are present in 2009, find the median individual among these 5 workers according to their 2005 wages, and then return the 2005 wages for this median individual as the “medindiv\_stayer\_appyr\_wages” variable.
* cht\_emps
  + The number of members of the application year cohort for the firm
* med\_cht\_wages
  + The **\*\*CONTEMPORANEOUS\*\*** median wages of application year cohort members in the firm-year grouping. Note that this retains those unemployed (i.e., 0 wages in the IRS data). If there are fewer than 3 stayers in this grouping, median application year cohort wages are set to missing.
* medindiv\_cht\_wages
  + The median wages of the application cohort employee in the associated year from the given firm-year grouping who was the median worker in the application year. If there are multiple individuals (due to a tie in median wages, or even number of workers) then this variable contains the mean of the application year cohort members’ wages in the associated year. If there are fewer than 3 cohort members in this grouping, median application year cohort wages are set to missing.
    - For example, say there are 10 members of the application year cohort for firm 20, who applied for a patent in 2004. The firm 20, year 2010 version of this variable would report the mean 2010 wages for the two workers who were ranks 5 and 6 in the wage ordering in the application year of 2004.
* medindiv\_cht\_appyr\_wages
  + The median application year wages of the application cohort employee in the associated year from the given firm-year grouping who was the median worker in the application year. If there are multiple individuals (due to a tie in median wages, or even number of workers) then this variable contains the mean of the application year cohort members’ wages in the application year. If there are fewer than 3 cohort members in this grouping, median wages for this variable are set to missing.
  + Note that this is just the same value of “medindiv\_cht\_wages” in the observation for the application year, but is provided in the output data set for easy differencing across years
    - For example, say there are 10 members of the application year cohort for firm 20, who applied for a patent in 2004. The firm 20, year 2010 version of this variable would report the mean 2004 wages for the two workers who were ranks 5 and 6 in the wage ordering in the application year of 2004.