**Questions**:

1. How to find the recode criterion they used for these data files?
   1. STURCTURETYPE
   2. FMTMETRO3
   3. REGION
2. Why numeric variable VACANCY has value ‘B’ in the original data file?

**Check points:**

1. Check how many observations has value of BURDEN greater than 1 (which means they are loaning for the house)

**Questions**:

1. What is the reasonable max value for BURDEN?
2. Should I exclude those observations that have problematic values for some variables?
   1. -6 value in VACANCY (91.78%), VALUE (100%) and PER (8.22%)
   2. -9 and -1 value in BURDEN
   3. 0 value in BEDRMS (0.91%)
   4. -9 value in AGE 1 (4033)
   5. 93 value in AGE 1 (429, 100%)
3. In general, how complex does this project suppose to be?

**Questions**:

1. The behavior of merge statement
   1. If I want to use the merge statement in data step to merge two datasets, but both of them contain observations that has the same ID, but different values for other variables (Both datasets have the same variables), what will the merge statement do if I merge two datasets by ID?
      1. Will SAS just remove that observation completely from the output dataset?
      2. How can I keep the observation I want in the output dataset? (For example, I want the observation that has larger values in variable A)

**Subset data:**

CONTROL (Control number)

VACANCY (Vacancy status)

BEDRMS (# of bedrooms in unit)

ROOMS (# of rooms in unit)

NUNITS (# of units in building)

BUILT (Year unit was built)

**STRUCTURETYPE** (Recoded structure type)

BURDEN (Housing cost as a fraction of income)

AGE1 (Age of head of household)

ZINC2 (Household Income)

ZSMHC (Monthly housing costs)

VALUE (Current market value of unit)

UTILITY (Monthly utility cost)

PER (# of persons in household)

**METRO3** (Central city / suburban status)

REGION (Census region)

**FMTMETRO3**

**1 - Large City:**

A central city of a CMSA or MSA, with the city having a population greater than or equal to 250,000.

**2 - Mid-size City:**

A central city of a CMSA or MSA, with the city having a population less than 250,000.

**3 - Urban Fringe of a Large City:**

Any territory within a CMSA or MSA of a Large City and defined as urban by the Census Bureau.

**4 - Urban Fringe of a Mid-size City:**

Any territory within a CMSA or MSA of a Mid-size City and defined as urban by the Census Bureau.

**5 - Large Town:**

An incorporated place or Census-designated place with a population greater than or equal to 25,000 and located outside a CMSA or MSA.

**6 - Small Town:**

An incorporated place or Census-designated place with a population less than 25,000 and greater than or equal to 2,500 and located outside a CMSA or MSA.

**7 - Rural, Outside MSA:**

Any territory designated as rural by the Census Bureau that is outside a CMSA or MSA of a Large or Mid-size City.

**8 - Rural, Inside MSA:**

Any territory designated as rural by the Census Bureau that is within a CMSA or MSA of a Large or Mid-size City.

**REGION**

**1 - Northeast**

**2 - Midwest**

**3 - South**

**4 - West**

**STRUCTURETYPE**

**1 = '1 Single Family'**

**2 = '2 2-4 units'**

**3 = '3 5-19 units'**

**4 = '4 20-49 units'**

**5 = '5 50+ units'**

**6 = '6 Mobile Home'**

**BURDEN**

For Q4 2012:

Mortgage: 12.6% (Average), should not exceed 30% (Los Angeles: 29%, San Francisco: 28.8%)

Burden: should not exceed 7 (Los Angeles and San Francisco: 6.8)

Reference: <http://www.forbes.com/sites/zillow/2013/04/16/high-home-price-to-income-ratios-hiding-behind-low-mortgage-rates/#2230f74d378d>