Capstone Adult Data Set (UCI)

Data Wrangling Steps

1. Review dataset description UCI website
2. Review data structure of input data via lapply
3. Test for missing and invalid values including outliers
4. Experiment plotting variables
5. Pre-Processing
   1. Centering
   2. Scaling
   3. Imputation
   4. Prediction
6. Create dummy variables
   1. Binary Variable income\_over\_50K – 0 if no 1 if yes from factor column income50 (<=50K, >50K)
7. Test correlation between variables to the dummy variable specifying income over 50k