(This exercise has been adapted from a data exercise from Empirical Analysis III course, taught by Prof. Mogstad and Prof. Heckman)

LaLonde (AER, 1986) investigated whether non-experimental methods could reproduce the experimental estimate based on the National Supported Work (NSW) Demonstration.

Dataset from Smith and Todd (J Ectrics, 2005) has been split into pieces that you will have to put together to effectively carry out the exercise.
The pieces include the NSW sample, as well as data from two non-experimental samples: one based on the Current Population Survey (CPS) and one on the Michigan Panel of Income Dynamics (PSID).

The variable -sample- identifies the relevant observations. The variable -treated- identifies the observations that were treated (participate in a subsidized work experience program) in the NSW (from April 1975 to August 1977).

You are interested in the average effect on Real Earnings in 1978 of the treatment for the treated.

Start with the NSW sample:

- Investigate whether the data is consistent with randomization of the treatment.
- Estimate the effect using the experimental sample.

Now use the sample consisting in the treated from the NSW sample and the comparison individuals from the CPS sample.

- Estimate the effect using OLS.
- Investigate covariate balancing and support between the treated and the CPS sample.
- (Bonus) Estimate the effect using 1 nearest neighbor propensity score matching. (Use -psmatch2- which can be installed using: ssc install psmatch2).