

```
options pagesize=50 linesize=80;
options formchar='|----|+|---+=|/\\<>*';
title1 "SYC Data Preprocessing";
footnote "SYC Data Preprocessing";

/* creating dataset */
data SYC;
  infile "/home/u63744989/dataset/syc.txt" missover firstobs = 2 delimiter = ',';
  /* before read in the data, we manually delete the instructions on the top of the txt file */
  input stratum psu psusize initwt finalwt randgrp age race ethnicity educ
  sex livewith famtime crimtype everviol numarr probtn corrinst evertime
  prviol prprop prdrug prpub prjuv agefirst usewepn alcuse everdrug;

  /* creating labels */
  label stratum = "stratum";
  label numarr = "number of prior arrest";

  /* Filter out rows with missing numarr value*/
  if (numarr = 99) then delete;

  /* Keep only the necessary columns, i.e. stratum & numarr*/
  keep stratum numarr;
run;

/* sort the dataset by stratum in an ascending order*/
proc sort data=SYC;
  by stratum;
run;

/* output dataset to a csv file*/
proc export data=SYC
  outfile='/home/u63744989/dataset/syc_post.csv'
  dbms=csv
  replace;
run;

/* View overview of the stratum variable */
proc freq data=SYC;
  tables stratum / nocol;
  title "Overview of Stratum Variable";
run;

ods pdf file='/home/u63744989/figures/output_1.pdf';

ods _all_ close;
```