Amanda Conley
Research Assignment #3
March 4, 2020

1.
    a. The DV is incwelfr. This is income whose source is through welfare benefits. According the IPUMS USA website: "INCWELFR reports how much pre-tax income (if any) the respondent received during the previous year from various public assistance programs commonly referred to as "welfare." Assistance from private charities was not included. The censuses collected information on income received from these sources during the previous calendar year; for the ACS and the PRCS, the reference period was the past 12 months."
    b. I am most interested in American Indian identity and health coverage status.
    c. The CVs are age (continuous) and sex (nominal).
    d. There is a statistically significant relationship between American Indian identity and income whose source is welfare, controlling for age and sex.
       There is a statistically significant relationship between American Indian identity, health coverage status and income whose source is welfare, controlling for age and sex.

2.
```
set more off
capture log close
cd "D:\Amanda's Files\UC Merced\Spring 2020\Soc211\Week 6"
log using RA3.log, replace
use incwelfr age sex racamind hcovany using usa_00002_california_only.dta
describe
summarize incwelfr age sex racamind hcovany

Contains data from usa_00002_california_only.dta
  obs:    1,872,509
 vars:          5                              2 Mar 2020 17:54
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
sex             byte    %8.0g      SEX        sex
age             byte    %8.0g      AGE        age
racamind        byte    %8.0g      RACAMIND   race: american indian or alaska
native
hcovany         byte    %8.0g      HCOVANY    any health insurance coverage
incwelfr        long    %12.0g                welfare (public assistance)
income
```

```
----------------------------------------------------------------------
Sorted by:

. summarize incwelfr age sex racamind hcovany

    Variable |        Obs        Mean    Std. Dev.        Min        Max
-------------+----------------------------------------------------------
    incwelfr |  1,872,509    17617.07    38025.68          0      99999
         age |  1,872,509      39.528    23.10361          0         94
         sex |  1,872,509    1.506706    .4999552          1          2
    racamind |  1,872,509    1.021118    .1437784          1          2
     hcovany |  1,872,509    1.905496     .292529          1          2
```

```
3. foreach x of var age sex racamind hcovany {
   quietly graph twoway (scatter incwelfr `x') (lfit incwelfr `x', color(red)),
  name(`x', replace) legend(off) ytitle(income) scheme(plotplainblind)
   }
  graph combine age sex racamind hcovany, col(3)
```
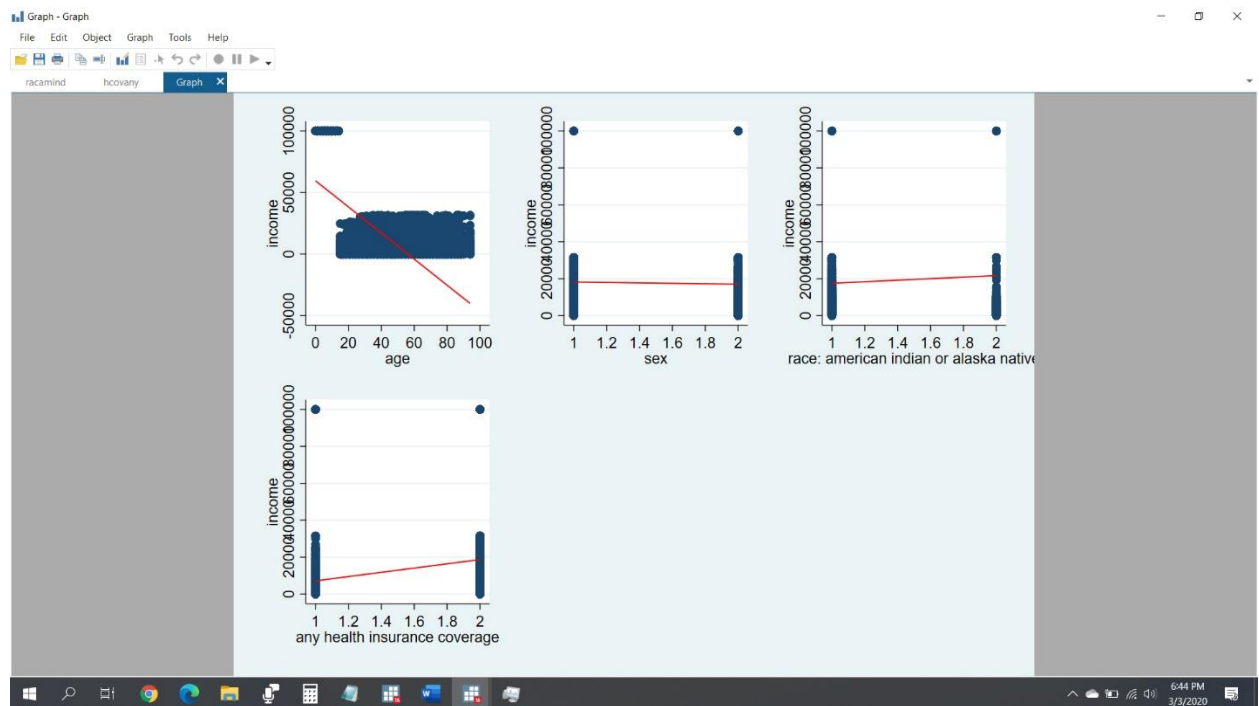

(the graphs produced EMFs that were so large that they kept crashing Word when I tried to add them. So, I took a screenshot instead.)

```
4.  correlate incwelfr age sex racamind hcovany

    pwcorr incwelfr age sex racamind hcovany, sig

    . correlate incwelfr age sex racamind hcovany
    (obs=1,872,509)

                 | incwelfr      age      sex racamind  hcovany
    -------------+------------------------------------------------
        incwelfr |   1.0000
             age |  -0.6440    1.0000
             sex |  -0.0157    0.0469    1.0000
        racamind |   0.0157   -0.0227   -0.0031    1.0000
         hcovany |   0.0886    0.0478    0.0451   -0.0184    1.0000


    .
    . pwcorr incwelfr age sex racamind hcovany, sig

                 | incwelfr      age      sex racamind  hcovany
    -------------+------------------------------------------------
        incwelfr |   1.0000
                 |
                 |
             age |  -0.6440    1.0000
                 |   0.0000
                 |
             sex |  -0.0157    0.0469    1.0000
                 |   0.0000    0.0000
                 |
        racamind |   0.0157   -0.0227   -0.0031    1.0000
                 |   0.0000    0.0000    0.0000
                 |
         hcovany |   0.0886    0.0478    0.0451   -0.0184    1.0000
                 |   0.0000    0.0000    0.0000    0.0000
                 |
```

I need to work with these variables a little bit more to feel comfortable with these results. However, for age, there is a strong negative relationship with incwelfr. Sex, American Indian racial identity, and health coverage status are all very weak or nonexistent positive relationships—I need to work on recoding these variables to be better suited for analysis. From a previous assignment, I also know that age interacts with health coverage status (age 19-26 produce a strong effect).

5.
   a. `reg incwelfr age sex racamind`
       The model above tests to see if there is an association between income whose source is welfare and American Indian identity, controlling for sex and age.
       `reg incwelfr age sex racamind hcovany`

The model above tests to see if there is an association between income whose source is welfare, American Indian identity and health coverage status, controlling for age and sex.

b. I believe, moving forward, adding a variable that includes whether or not the respondent resides in an area considered by the census to be tribal lands might intervene, strengthening the relationship between income whose source is welfare and American Indian identity. I also think it would intervene in health coverage status.

6.

    a. $incwelfr = \alpha + \beta_1.age_1 + \beta_2.sex_2 + \beta_3.racamind_3 \cdot \chi_3$

    b. $incwelfr = \alpha + \beta_1.age_1 + \beta_2.sex_2 + \beta_3.racamind_3 + \beta_4.hcovany_4 \cdot \chi_4$

7. I used estout in STATA to produce a .rtf file of this table.

	Base Model	Model 2
Age	-1.1e+03***	-1.1e+03***
	(0.921)	(0.911)
Male	0.000	0.000
	(.)	(.)
Female	1104.017***	714.971***
	(42.561)	(42.082)
No	0.000	0.000
	(.)	(.)
Yes	302.859*	845.784***
	(147.872)	(146.093)
No health insurance coverage		0.000
		(.)
With health insurance coverage		1.6e+04***
		(71.935)
Constant	5.9e+04***	4.5e+04***
	(46.739)	(77.817)
R^2	0.415	0.429
Adjusted R^2	0.415	0.429
BIC	4.4e+07	4.4e+07
Observations	1872509	1872509

8.

a. I don't think either model is rejected, though the second model explains more than the base model.

b. I suspect that some of the IVs may be too highly correlated or need to be recoded.

c. I think this analysis needs to explore location as a source of explanation for income whose source is welfare.