---

title: "Homework#4"

author: "Neshma Simon"

date: "10/14/2020"

output: html\_document

---

## Homework 4

## Group Members: Fareha, Hertz, and Stan

```

load("acs2017\_ny\_data.Rdata")

attach(acs2017\_ny)

```

>For this analysis we will be using the subgroup of people whose ages are in the range of 25 to 55. This subgroup is ideal because this group is most likely part of the labor force and work full time. This allows us to exclude people who are unemployed with high qualifications. Using the age range between 25-55 ensures that this range is fit for people of the working full-time age.

```

use\_varb <- (AGE >= 25) & (AGE <= 55) & (LABFORCE == 2) & (WKSWORK2 > 4) & (UHRSWORK >= 35)

dat\_use <- subset(acs2017\_ny,use\_varb) #

summary (dat\_use)

AGE female educ\_nohs educ\_hs educ\_somecoll

Min. :25.0 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.00

1st Qu.:33.0 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.00

Median :41.0 Median :0.0000 Median :0.0000 Median :0.0000 Median :0.00

Mean :40.5 Mean :0.4526 Mean :0.0466 Mean :0.2579 Mean :0.21

3rd Qu.:49.0 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:0.00

Max. :55.0 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.00

educ\_college educ\_advdeg SCHOOL

Min. :0.0000 Min. :0.0000 N/A : 0

1st Qu.:0.0000 1st Qu.:0.0000 No, not in school:44768

Median :0.0000 Median :0.0000 Yes, in school : 2203

Mean :0.2775 Mean :0.2079 Missing : 0

3rd Qu.:1.0000 3rd Qu.:0.0000

Max. :1.0000 Max. :1.0000

EDUC EDUCD

4 years of college :13035 Bachelor's degree :13035

Grade 12 :12116 Regular high school diploma : 7768

5+ years of college: 9766 Master's degree : 7069

2 years of college : 5013 Associate's degree, type not specified : 5013

1 year of college : 4852 1 or more years of college credit, no degree: 4852

Grade 5, 6, 7, or 8: 617 Some college, but less than 1 year : 2210

(Other) : 1572 (Other) : 7024

DEGFIELD

N/A :24170

Business : 4836

Social Sciences : 2178

Education Administration and Teaching : 1859

Medical and Health Sciences and Services: 1481

Fine Arts : 1432

(Other) :11015

DEGFIELDD

N/A :24170

Psychology : 1236

Business Management and Administration: 1188

Accounting : 1011

General Business : 873

English Language and Literature : 805

(Other) :17688

DEGFIELD2

N/A :44103

Business : 495

Social Sciences : 426

Fine Arts : 205

Education Administration and Teaching: 192

Communications : 180

(Other) : 1370

DEGFIELD2D

N/A :44103

Economics : 136

Political Science and Government : 124

Psychology : 112

Business Management and Administration : 110

French, German, Latin and Other Common Foreign Language Studies: 101

(Other) : 2285

PUMA GQ OWNERSHP OWNERSHPD MORTGAGE

Min. : 100 Min. :1.000 Min. :0.000 Min. : 0.00 Min. :0.000

1st Qu.:1801 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:13.00 1st Qu.:0.000

Median :3207 Median :1.000 Median :1.000 Median :13.00 Median :3.000

Mean :2809 Mean :1.017 Mean :1.353 Mean :16.01 Mean :1.646

3rd Qu.:4003 3rd Qu.:1.000 3rd Qu.:2.000 3rd Qu.:22.00 3rd Qu.:3.000

Max. :4114 Max. :5.000 Max. :2.000 Max. :22.00 Max. :4.000

OWNCOST RENT COSTELEC COSTGAS COSTWATR

Min. : 0 Min. : 0.0 Min. : 0 Min. : 0 Min. : 0

1st Qu.: 1415 1st Qu.: 0.0 1st Qu.: 960 1st Qu.: 960 1st Qu.: 400

Median : 3032 Median : 0.0 Median :1560 Median :3120 Median :3300

Mean :37479 Mean : 527.1 Mean :2288 Mean :5326 Mean :5244

3rd Qu.:99999 3rd Qu.:1000.0 3rd Qu.:2400 3rd Qu.:9993 3rd Qu.:9995

Max. :99999 Max. :3800.0 Max. :9997 Max. :9997 Max. :9997

COSTFUEL HHINCOME FOODSTMP LINGISOL ROOMS

Min. : 0 Min. : -11800 Min. :1.000 Min. :0.000 Min. : 0.000

1st Qu.:9993 1st Qu.: 67300 1st Qu.:1.000 1st Qu.:1.000 1st Qu.: 4.000

Median :9993 Median : 109000 Median :1.000 Median :1.000 Median : 6.000

Mean :8479 Mean : 141457 Mean :1.071 Mean :1.044 Mean : 6.021

3rd Qu.:9993 3rd Qu.: 168500 3rd Qu.:1.000 3rd Qu.:1.000 3rd Qu.: 8.000

Max. :9997 Max. :2030000 Max. :2.000 Max. :2.000 Max. :16.000

NA's :203

BUILTYR2 UNITSSTR FUELHEAT SSMC

Min. : 0.000 Min. : 0.00 Min. :0.000 Min. :0.00000

1st Qu.: 1.000 1st Qu.: 3.00 1st Qu.:2.000 1st Qu.:0.00000

Median : 3.000 Median : 3.00 Median :2.000 Median :0.00000

Mean : 4.009 Mean : 4.76 Mean :3.068 Mean :0.01618

3rd Qu.: 5.000 3rd Qu.: 6.00 3rd Qu.:4.000 3rd Qu.:0.00000

Max. :22.000 Max. :10.00 Max. :9.000 Max. :2.00000

FAMSIZE NCHILD NCHLT5 RELATE

Min. : 1.000 Min. :0.0000 Min. :0.0000 Min. : 1.000

1st Qu.: 2.000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.: 1.000

Median : 3.000 Median :1.0000 Median :0.0000 Median : 1.000

Mean : 3.096 Mean :0.9721 Mean :0.1986 Mean : 2.624

3rd Qu.: 4.000 3rd Qu.:2.0000 3rd Qu.:0.0000 3rd Qu.: 2.000

Max. :19.000 Max. :9.0000 Max. :5.0000 Max. :12.000

RELATED MARST RACE RACED HISPAN

Min. : 101.0 Min. :1.000 Min. :1.00 Min. :100 Min. :0.0000

1st Qu.: 101.0 1st Qu.:1.000 1st Qu.:1.00 1st Qu.:100 1st Qu.:0.0000

Median : 101.0 Median :1.000 Median :1.00 Median :100 Median :0.0000

Mean : 265.6 Mean :2.931 Mean :2.05 Mean :207 Mean :0.4283

3rd Qu.: 201.0 3rd Qu.:6.000 3rd Qu.:2.00 3rd Qu.:200 3rd Qu.:0.0000

Max. :1270.0 Max. :6.000 Max. :9.00 Max. :990 Max. :4.0000

HISPAND BPL BPLD

Min. : 0 New York :27217 New York :27217

1st Qu.: 0 West Indies : 2521 China : 1192

Median : 0 SOUTH AMERICA: 1646 Dominican Republic: 1015

Mean : 46 China : 1494 New Jersey : 893

3rd Qu.: 0 India : 1048 Pennsylvania : 800

Max. :498 New Jersey : 893 Jamaica : 651

(Other) :12152 (Other) :15203

ANCESTR1 ANCESTR1D

Not Reported : 7125 Not Reported : 7125

Italian : 5207 Italian (1990-2000, ACS, PRCS) : 5207

Irish, various subheads,: 4039 Irish : 3842

German : 3207 German (1990-2000, ACS/PRCS) : 3182

African-American : 1792 African-American (1990-2000, ACS, PRCS): 1792

Polish : 1641 Polish : 1641

(Other) :23960 (Other) :24182

ANCESTR2 ANCESTR2D CITIZEN

Not Reported:33289 Not Reported :33289 Min. :0.0000

Irish : 2487 German (1990-2000, ACS, PRCS) : 2398 1st Qu.:0.0000

German : 2403 Irish : 2380 Median :0.0000

English : 1181 English : 1181 Mean :0.6219

Italian : 1084 Italian (1990-2000, ACS, PRCS): 1084 3rd Qu.:2.0000

Polish : 886 Polish : 886 Max. :3.0000

(Other) : 5641 (Other) : 5753

YRSUSA1 HCOVANY HCOVPRIV SEX EMPSTAT

Min. : 0.00 Min. :1.000 Min. :1.000 Male :25713 Min. :1.000

1st Qu.: 0.00 1st Qu.:2.000 1st Qu.:2.000 Female:21258 1st Qu.:1.000

Median : 0.00 Median :2.000 Median :2.000 Median :1.000

Mean : 5.47 Mean :1.942 Mean :1.862 Mean :1.003

3rd Qu.: 4.00 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:1.000

Max. :56.00 Max. :2.000 Max. :2.000 Max. :2.000

EMPSTATD LABFORCE OCC IND CLASSWKR

Min. :10.00 Min. :2 430 : 1641 7860 : 3427 Min. :1.000

1st Qu.:10.00 1st Qu.:2 2310 : 1628 8190 : 2940 1st Qu.:2.000

Median :10.00 Median :2 5700 : 1105 770 : 2780 Median :2.000

Mean :10.07 Mean :2 4700 : 1077 8680 : 1663 Mean :1.921

3rd Qu.:10.00 3rd Qu.:2 3255 : 1036 9470 : 1548 3rd Qu.:2.000

Max. :20.00 Max. :2 800 : 932 7870 : 1404 Max. :2.000

(Other):39552 (Other):33209

CLASSWKRD WKSWORK2 UHRSWORK INCTOT

Min. :13.00 Min. :5.000 Min. :35.00 Min. : -5900

1st Qu.:22.00 1st Qu.:6.000 1st Qu.:40.00 1st Qu.: 35000

Median :22.00 Median :6.000 Median :40.00 Median : 55000

Mean :22.43 Mean :5.978 Mean :43.91 Mean : 77282

3rd Qu.:23.00 3rd Qu.:6.000 3rd Qu.:47.00 3rd Qu.: 88000

Max. :29.00 Max. :6.000 Max. :99.00 Max. :1378000

FTOTINC INCWAGE POVERTY MIGRATE1

Min. : -11800 Min. : 0 Min. : 0.0 Min. :1.000

1st Qu.: 55000 1st Qu.: 32000 1st Qu.:307.0 1st Qu.:1.000

Median : 96000 Median : 52000 Median :491.0 Median :1.000

Mean : 128560 Mean : 72477 Mean :400.2 Mean :1.132

3rd Qu.: 155280 3rd Qu.: 85000 3rd Qu.:501.0 3rd Qu.:1.000

Max. :2030000 Max. :638000 Max. :501.0 Max. :4.000

NA's :203

MIGRATE1D MIGPLAC1 MIGCOUNTY1 MIGPUMA1

Min. :10.00 Min. : 0.00 Min. : 0.0 Min. : 0.0

1st Qu.:10.00 1st Qu.: 0.00 1st Qu.: 0.0 1st Qu.: 0.0

Median :10.00 Median : 0.00 Median : 0.0 Median : 0.0

Mean :11.65 Mean : 5.43 Mean : 4.9 Mean : 325.2

3rd Qu.:10.00 3rd Qu.: 0.00 3rd Qu.: 0.0 3rd Qu.: 0.0

Max. :40.00 Max. :900.00 Max. :810.0 Max. :70100.0

VETSTAT VETSTATD PWPUMA00 TRANWORK TRANTIME

Min. :1.000 Min. :11.00 Min. : 0 Min. : 0.00 Min. : 0.00

1st Qu.:1.000 1st Qu.:11.00 1st Qu.: 1300 1st Qu.:10.00 1st Qu.: 15.00

Median :1.000 Median :11.00 Median : 3200 Median :10.00 Median : 30.00

Mean :1.029 Mean :11.27 Mean : 2716 Mean :20.15 Mean : 33.37

3rd Qu.:1.000 3rd Qu.:11.00 3rd Qu.: 3800 3rd Qu.:33.00 3rd Qu.: 45.00

Max. :2.000 Max. :20.00 Max. :59300 Max. :70.00 Max. :138.00

DEPARTS in\_NYC in\_Bronx in\_Manhattan

Min. : 0.0 Min. :0.0000 Min. :0.00000 Min. :0.00000

1st Qu.: 632.0 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.00000

Median : 732.0 Median :0.0000 Median :0.00000 Median :0.00000

Mean : 770.7 Mean :0.3989 Mean :0.04545 Mean :0.06402

3rd Qu.: 832.0 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:0.00000

Max. :2345.0 Max. :1.0000 Max. :1.00000 Max. :1.00000

in\_StatenI in\_Brooklyn in\_Queens in\_Westchester

Min. :0.00000 Min. :0.0000 Min. :0.0000 Min. :0.0000

1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000

Median :0.00000 Median :0.0000 Median :0.0000 Median :0.0000

Mean :0.02131 Mean :0.1406 Mean :0.1275 Mean :0.0449

3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:0.0000

Max. :1.00000 Max. :1.0000 Max. :1.0000 Max. :1.0000

in\_Nassau Hispanic Hisp\_Mex Hisp\_PR

Min. :0.00000 Min. :0.000 Min. :0.00000 Min. :0.00000

1st Qu.:0.00000 1st Qu.:0.000 1st Qu.:0.00000 1st Qu.:0.00000

Median :0.00000 Median :0.000 Median :0.00000 Median :0.00000

Mean :0.07011 Mean :0.141 Mean :0.01829 Mean :0.03879

3rd Qu.:0.00000 3rd Qu.:0.000 3rd Qu.:0.00000 3rd Qu.:0.00000

Max. :1.00000 Max. :1.000 Max. :1.00000 Max. :1.00000

Hisp\_Cuban Hisp\_DomR white AfAm

Min. :0.0000 Min. :0.00000 Min. :0.0000 Min. :0.0000

1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.0000

Median :0.0000 Median :0.00000 Median :1.0000 Median :0.0000

Mean :0.0033 Mean :0.02793 Mean :0.6992 Mean :0.1159

3rd Qu.:0.0000 3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:0.0000

Max. :1.0000 Max. :1.00000 Max. :1.0000 Max. :1.0000

Amindian Asian race\_oth unmarried

Min. :0.000000 Min. :0.00000 Min. :0.0000 Min. :0.0000

1st Qu.:0.000000 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.0000

Median :0.000000 Median :0.00000 Median :0.0000 Median :0.0000

Mean :0.003428 Mean :0.09895 Mean :0.1364 Mean :0.3159

3rd Qu.:0.000000 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:1.0000

Max. :1.000000 Max. :1.00000 Max. :1.0000 Max. :1.0000

veteran has\_AnyHealthIns has\_PvtHealthIns Commute\_car

Min. :0.00000 Min. :0.0000 Min. :0.0000 Min. :0.0000

1st Qu.:0.00000 1st Qu.:1.0000 1st Qu.:1.0000 1st Qu.:0.0000

Median :0.00000 Median :1.0000 Median :1.0000 Median :1.0000

Mean :0.02859 Mean :0.9421 Mean :0.8624 Mean :0.6306

3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000

Max. :1.00000 Max. :1.0000 Max. :1.0000 Max. :1.0000

Commute\_bus Commute\_subway Commute\_rail Commute\_other

Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.00000

1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.00000

Median :0.0000 Median :0.0000 Median :0.0000 Median :0.00000

Mean :0.0377 Mean :0.1845 Mean :0.0344 Mean :0.09576

3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:0.00000

Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.00000

below\_povertyline below\_150poverty below\_200poverty foodstamps

Min. :0.00000 Min. :0.00000 Min. :0.0000 Min. :0.00000

1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.00000

Median :0.00000 Median :0.00000 Median :0.0000 Median :0.00000

Mean :0.02169 Mean :0.06119 Mean :0.1167 Mean :0.07083

3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:0.00000

Max. :1.00000 Max. :1.00000 Max. :1.0000 Max. :1.00000

detach()

attach(dat\_use)

```

>In this range of 25-55, we wanted to see what factors can influence income wages so we ran a model test to see the results. Then, we try linear regression with the dat we have. In this, we set the wage as dependent and a dummy.

```

model\_temp1 <- lm(INCWAGE ~ AGE + female + AfAm + Asian + Amindian + race\_oth + Hispanic + educ\_hs + educ\_somecoll + educ\_college + educ\_advdeg)

summary(model\_temp1)

```

>The linear regression gives us the following data:

```

Call:

lm(formula = INCWAGE ~ AGE + female + AfAm + Asian + Amindian +

race\_oth + Hispanic + educ\_hs + educ\_somecoll + educ\_college +

educ\_advdeg)

Residuals:

Min 1Q Median 3Q Max

-148088 -33205 -10708 13053 625543

```

>We see here that there is a median, -10708, of a negative value.

```

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -7096.25 2446.71 -2.900 0.003730 \*\*

AGE 1316.69 39.66 33.199 < 2e-16 \*\*\*

female -24939.46 720.43 -34.617 < 2e-16 \*\*\*

AfAm -11934.26 1130.37 -10.558 < 2e-16 \*\*\*

Asian 566.53 1369.83 0.414 0.679188

Amindian -8858.57 6077.71 -1.458 0.144971

race\_oth -7526.49 1272.49 -5.915 3.35e-09 \*\*\*

Hispanic -4224.82 1183.47 -3.570 0.000358 \*\*\*

educ\_hs 10592.37 1814.71 5.837 5.35e-09 \*\*\*

educ\_somecoll 22461.39 1857.67 12.091 < 2e-16 \*\*\*

educ\_college 57155.71 1830.96 31.216 < 2e-16 \*\*\*

educ\_advdeg 82766.43 1878.64 44.057 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 76760 on 46959 degrees of freedom

Multiple R-squared: 0.15, Adjusted R-squared: 0.1498

F-statistic: 753.6 on 11 and 46959 DF, p-value: < 2.2e-16

```

```

par(mfrow=c(2,2))

plot(model\_temp1,col="red",pch=16,cex=1,lwd=1,lty=2)

## view: PDF titled "plot1"

```

```

require(stargazer)

stargazer(model\_temp1, type = "text")

===============================================

Dependent variable:

---------------------------

INCWAGE

-----------------------------------------------

AGE 1,316.691\*\*\*

(39.661)

female -24,939.460\*\*\*

(720.433)

AfAm -11,934.250\*\*\*

(1,130.372)

Asian 566.528

(1,369.834)

Amindian -8,858.569

(6,077.710)

race\_oth -7,526.487\*\*\*

(1,272.485)

Hispanic -4,224.816\*\*\*

(1,183.469)

educ\_hs 10,592.370\*\*\*

(1,814.709)

educ\_somecoll 22,461.390\*\*\*

(1,857.674)

educ\_college 57,155.710\*\*\*

(1,830.963)

educ\_advdeg 82,766.430\*\*\*

(1,878.638)

Constant -7,096.252\*\*\*

(2,446.712)

-----------------------------------------------

Observations 46,971

R2 0.150

Adjusted R2 0.150

Residual Std. Error 76,755.980 (df = 46959)

F Statistic 753.551\*\*\* (df = 11; 46959)

===============================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

```

```

nAmindian<-as.numeric(as.character(dat\_use$INCWAGE))

par(mfrow=c(2,2))

Wage\_Amindian<-lm(INCWAGE~Amindian)

plot(Wage\_Amindian,col="green",pch=14,cex=1,lwd=1,lty=2)

## view: PDF titled "plot2"

summary(Wage\_Amindian)

Call:

lm(formula = INCWAGE ~ Amindian)

Residuals:

Min 1Q Median 3Q Max

-72553 -40553 -20553 12447 587481

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 72552.5 384.7 188.586 <2e-16 \*\*\*

Amindian -22033.3 6571.2 -3.353 8e-04 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 83240 on 46969 degrees of freedom

Multiple R-squared: 0.0002393, Adjusted R-squared: 0.000218

F-statistic: 11.24 on 1 and 46969 DF, p-value: 8e-04

```

>Based on the summary of the model temp, we can see that the p-value of Amindian is 0.144971. To us, this can be deemed as significant, which meant we needed to create an additional variable to see if these two can create another significant regression coefficient.

>The overall multiple R-squared is 0.15 and the adjusted R-squared is 0.1498.

>We decided to use the additional variable of Hispanics to see their correlation to Amindian with the use of INCWAGE.

```

nHispanic<-as.numeric(as.character(dat\_use$INCWAGE))

par(mfrow=c(2,2))

Wage\_Hispanic<-lm(INCWAGE~Hispanic)

plot(Wage\_Hispanic,col="purple",pch=14,cex=1,lwd=1,lty=2)

summary(Wage\_Hispanic)

## view: PDF titled "plot3"

summary(Wage\_Hispanic)

all:

lm(formula = INCWAGE ~ Hispanic)

Residuals:

Min 1Q Median 3Q Max

-75702 -39702 -18702 12168 585168

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 75701.7 412.5 183.50 <2e-16 \*\*\*

Hispanic -22869.5 1098.6 -20.82 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 82860 on 46969 degrees of freedom

Multiple R-squared: 0.009142, Adjusted R-squared: 0.00912

F-statistic: 433.3 on 1 and 46969 DF, p-value: < 2.2e-16

```

> Based on the summary of the model temp, we can see that the p-value of Hispanic is 0.000358. We chose these two variables of Amindian and Hispanic due to their smaller p-value as compared to the other vairables. To us, these regression coefficients can be deemed as statistically significant.

>The last thing we did was use this code to get a regression line in a plot with all the data points to give us an idea about the relationship between the dependent variables and the independent variable we performed the regression on. It shows us how one variable changes due to a change in the other.

```

require(AER)

NNobs <- length(INCWAGE)

set.seed(12345)

graph\_obs <- (runif(NNobs) < 0.1)

dat\_graph <-subset(dat\_use,graph\_obs)

```

> We tried the summary and plots of the new variables just to see the progression and differences in each output.

```

plot(INCWAGE ~ jitter(AGE, factor = 2), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), data = dat\_graph)

plot(INCWAGE ~ jitter(AGE, factor = 2), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), ylim = c(0,150000), data = dat\_graph)

to\_be\_predicted2 <- data.frame(AGE = 25:55, female = 1, AfAm = 0, Asian = 0, Amindian = 1, race\_oth = 1, Hispanic = 1, educ\_hs = 0, educ\_somecoll = 0, educ\_college = 1, educ\_advdeg = 0)

to\_be\_predicted2$yhat <- predict(model\_temp1, newdata = to\_be\_predicted2)

lines(yhat ~ AGE, data = to\_be\_predicted2)

## view: PDF titled "plot4"

```

```

NNobs <- length(INCWAGE)

set.seed(12345)

graph\_obs <- (runif(NNobs) < 0.1)

dat\_graph <-subset(dat\_use,graph\_obs)

plot(INCWAGE ~ jitter(AGE, factor = 2), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), ylim = c(0,150000), data = dat\_graph)

to\_be\_predicted2 <- data.frame(AGE = 25:55, female = 1, AfAm = 0, Asian = 0, Amindian = 1, race\_oth = 1, Hispanic = 1, educ\_hs = 0, educ\_somecoll = 0, educ\_college = 1, educ\_advdeg = 0)

to\_be\_predicted2$yhat <- predict(model\_temp1, newdata = to\_be\_predicted2)

lines(yhat ~ AGE, data = to\_be\_predicted2)

## view: PDF titled "plot5"

```

```

NNobs <- length(INCWAGE)

set.seed(12345)

graph\_obs <- (runif(NNobs) < 0.1)

dat\_graph <-subset(dat\_use,graph\_obs)

plot(INCWAGE ~ jitter(AGE, factor = 2), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), ylim = c(0,150000), data = dat\_graph)

to\_be\_predicted2 <- data.frame(AGE = 25:55, female = 1, AfAm = 0, Asian = 0, Amindian = 1, race\_oth = 1, Hispanic = 1, educ\_hs = 0, educ\_somecoll = 0, educ\_college = 1, educ\_advdeg = 0)

to\_be\_predicted2$yhat <- predict(model\_temp1, newdata = to\_be\_predicted2)

lines(yhat ~ AGE, data = to\_be\_predicted2)

NNNobs <- length(INCTOT)

set.seed(12345)

graph\_obs <- (runif(NNobs) < 0.1)

dat\_graph <-subset(dat\_use,graph\_obs)

plot(INCTOT ~ jitter(AGE, factor = 2), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), ylim = c(0,150000), data = dat\_graph)

to\_be\_predicted4 <- data.frame(AGE = 25:55, educ\_hs = 0, educ\_somecoll = 0)

to\_be\_predicted4$yhat <- predict(model\_temp1, newdata = to\_be\_predicted2)

lines(yhat ~ AGE, data = to\_be\_predicted2)

## view: PDF titled "plot6"

```

```

NNobs <- length(INCTOT)

set.seed(12345)

graph\_obs <- (runif(NNobs) < 0.1)

dat\_graph <-subset(dat\_use,graph\_obs)

plot(INCTOT ~ jitter(AGE, factor = 2), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), data = dat\_graph)

plot(INCTOT ~ jitter(AGE, factor = 2), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), ylim = c(0,150000), data = dat\_graph)

to\_be\_predicted2 <- data.frame(AGE = 25:55, female = 1, AfAm = 0, Asian = 0, Amindian = 1, race\_oth = 1, Hispanic = 1, educ\_hs = 0, educ\_somecoll = 0, educ\_college = 1, educ\_advdeg = 0)

to\_be\_predicted2$yhat <- predict(model\_temp1, newdata = to\_be\_predicted2)

lines(yhat ~ AGE, data = to\_be\_predicted2)

## view: PDF titled "plot7"

```

```

summary(to\_be\_predicted2)

AGE female AfAm Asian Amindian race\_oth

Min. :25.0 Min. :1 Min. :0 Min. :0 Min. :1 Min. :1

1st Qu.:32.5 1st Qu.:1 1st Qu.:0 1st Qu.:0 1st Qu.:1 1st Qu.:1

Median :40.0 Median :1 Median :0 Median :0 Median :1 Median :1

Mean :40.0 Mean :1 Mean :0 Mean :0 Mean :1 Mean :1

3rd Qu.:47.5 3rd Qu.:1 3rd Qu.:0 3rd Qu.:0 3rd Qu.:1 3rd Qu.:1

Max. :55.0 Max. :1 Max. :0 Max. :0 Max. :1 Max. :1

Hispanic educ\_hs educ\_somecoll educ\_college educ\_advdeg yhat

Min. :1 Min. :0 Min. :0 Min. :1 Min. :0 Min. :37427

1st Qu.:1 1st Qu.:0 1st Qu.:0 1st Qu.:1 1st Qu.:0 1st Qu.:47303

Median :1 Median :0 Median :0 Median :1 Median :0 Median :57178

Mean :1 Mean :0 Mean :0 Mean :1 Mean :0 Mean :57178

3rd Qu.:1 3rd Qu.:0 3rd Qu.:0 3rd Qu.:1 3rd Qu.:0 3rd Qu.:67053

Max. :1 Max. :0 Max. :0 Max. :1 Max. :0 Max. :76928

summary(INCTOT)

Min. 1st Qu. Median Mean 3rd Qu. Max.

-5900 35000 55000 77282 88000 1378000

summary(NNobs)

Min. 1st Qu. Median Mean 3rd Qu. Max.

46971 46971 46971 46971 46971 46971

```

>Through the plot that was created, with age and INCTOT, we see a positive sloping line which indicates a correlation between both variables.