Kerra Sinanan
Exam 2

Question 1:

In the first question we have two separate regressions

For White-HH (W):
$YW=\beta0+\beta1Age+\varepsilon$

For Non-White HH (NW):

$YNW=\alpha0+\alpha1Age+v$

Full pooled regression with dummy D (D = 1 if NW)
$Y=\gamma0+\gamma1Age+\gamma2D+\gamma3(D\times Age)+uY$

Interpretation:

- γ0: baseline intercept (for the reference group, D=0 → White)
- γ1: baseline Age slope (for White)
- γ2: *how much the intercept changes* when D goes from 0 to 1 (NW vs W)
- γ3: *how much the Age slope changes* when D goes from 0 to 1

For White HH, D=0D = 0D=0.

Start from the pooled model:

$Y=\gamma0+\gamma1Age+\gamma2D+\gamma3(D\times Age)+u$

Set D=0D = 0D=0:

$YW=\gamma0+\gamma1Age+\gamma2\cdot0+\gamma3\cdot0\cdot Age+u$

So this simplifies to:

$YW=\gamma0+\gamma1Age+uY_W = \gamma_0 + \gamma_1 \text{Age} + uYW=\gamma0+\gamma1Age+u$

Compare that to the original White regression:

$YW=\beta0+\beta1Age+\varepsilon$

- Constant term: $\beta_0 = \gamma_0$
- Age coefficient: $\beta_1 = \gamma_1$

So:

$\beta_0 = \gamma_0, \beta_1 = \gamma_1$- The intercept and slope for the White group are exactly the baseline intercept and baseline Age effect in the pooled model.

For Non-White HH, $D = 1$.

Start again from the pooled model:

$Y = \gamma_0 + \gamma_1 Age + \gamma_2 D + \gamma_3 (D \times Age) + u$

Set $D = 1$:

$Y_{NW} = \gamma_0 + \gamma_1 \text{Age} + \gamma_2 \cdot 1 + \gamma_3 \cdot 1 \cdot \text{Age} + u$

Simplify:

$Y_{NW} = (\gamma_0 + \gamma_2) + (\gamma_1 + \gamma_3) Age + u$

Compare to the original NW regression:

$Y_{NW} = \alpha_0 + \alpha_1 Age + v$

Match intercept and slope:

- Intercept: $\alpha_0 = \gamma_0 + \gamma_2$
- Age coefficient: $\alpha_1 = \gamma_1 + \gamma_3$

So:

$\alpha_0 = \gamma_0 + \gamma_2, \alpha_1 = \gamma_1 + \gamma_3$- The intercept and slope for the Non-White group are the baseline values plus the “extra” shifts from γ_2 and γ_3.

Let D = 1 if the household head is Non-White (NW) and 0 if White (W).
The pooled regression is
$Y = \gamma_0 + \gamma_1 Age + \gamma_2 D + \gamma_3 (D \cdot Age) + u$

When $D=0$ (White HH):
$YW=\gamma0+\gamma1Age+u$
Comparing with $YW=\beta0+\beta1Age+\varepsilon$

gives
$\beta0=\gamma0$, $\beta1=\gamma1$

When $D=1$ (Non-White HH):
$YNW=(\gamma0+\gamma2)+(\gamma1+\gamma3)Age+u$
Comparing with $YNW=\alpha0+\alpha1Age+v$
$\alpha0=\gamma0+\gamma2$, $\alpha1=\gamma1+\gamma3$

Therefore the relationships are:

$\beta0=\gamma0, \beta1=\gamma1, \alpha0=\gamma0+\gamma2, \alpha1=\gamma1+\gamma3$

and equivalently

$\gamma2=\alpha0-\beta0, \gamma3=\alpha1-\beta1$

Economically, $\gamma2$\gamma_2$\gamma2$ measures the difference in intercepts (baseline Y) between NW and W households, and $\gamma3$\gamma_3$\gamma3$ measures the difference in the effect of Age on Y between NW and W households.

Question 2:

```
> HHP <- Household_Pulse_data

>

> recode_anx <- c(

+    "no anxiety over past 2 wks"                          = 0,

+    "several days anxiety over past 2 wks"                = 1,

+    "more than half the days anxiety over past 2 wks"     = 2,

+    "nearly every day anxiety"                            = 3

+ )

>

> recode_worry <- c(

+    "no worry over past 2 wks"                            = 0,
```

```
+    "several days worried over past 2 wks"               = 1,
+    "more than half the days worried over past 2 wks"    = 2,
+    "nearly every day worry"                             = 3
+ )
>
> HHP$ANX_num   <- recode_anx[HHP$ANXIOUS]
> HHP$WORRY_num <- recode_worry[HHP$WORRY]
> HHP$K4SUM     <- HHP$ANX_num + HHP$WORRY_num
>
> # Drop missing values in key vars
> HHP2 <- subset(HHP, !is.na(K4SUM) & !is.na(EEDUC) & !is.na(INCOME))
>
> # Treat education and income as categorical
> HHP2$EEDUC  <- factor(HHP2$EEDUC)
> HHP2$INCOME <- factor(HHP2$INCOME)
>
> # Baseline OLS model
> m2 <- lm(K4SUM ~ EEDUC + INCOME, data = HHP2)
> summary(m2)


Call:
lm(formula = K4SUM ~ EEDUC + INCOME, data = HHP2)


Residuals:
    Min      1Q  Median      3Q     Max
-3.4803 -1.0838 -0.6868  0.9222  5.0459
```

```
Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      0.95406    0.07242  13.173  < 2e-16 ***
EEDUCsome hs                     0.19066    0.08634   2.208   0.0272 *
EEDUCHS diploma                  0.18902    0.07370   2.565   0.0103 *
EEDUCsome coll                   0.32332    0.07297   4.431 9.41e-06 ***
EEDUCassoc deg                   0.33428    0.07389   4.524 6.08e-06 ***
EEDUCbach deg                    0.35744    0.07286   4.906 9.31e-07 ***
EEDUCadv deg                     0.35549    0.07306   4.866 1.14e-06 ***
INCOMEHH income less than $25k   2.16877    0.02366  91.656  < 2e-16 ***
INCOMEHH income $25k - $34.9k    1.98938    0.02456  80.988  < 2e-16 ***
INCOMEHH income $35k - 49.9      1.94077    0.02231  87.004  < 2e-16 ***
INCOMEHH income $50k - 74.9      1.80039    0.01866  96.485  < 2e-16 ***
INCOMEHH income $75 - 99.9       1.73229    0.01966  88.104  < 2e-16 ***
INCOMEHH income $100k - 149      1.66574    0.01817  91.651  < 2e-16 ***
INCOMEHH income $150 - 199       1.58783    0.02309  68.753  < 2e-16 ***
INCOMEHH income $200k +          1.54374    0.02164  71.340  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.317 on 59825 degrees of freedom
Multiple R-squared:  0.2535,   Adjusted R-squared:  0.2533
F-statistic:  1451 on 14 and 59825 DF,  p-value: < 2.2e-16

>
```

```
> m2 <- lm(K4SUM ~ EEDUC + INCOME + ANYWORK + REGION + TENURE, data = HHP2)
> summary(m2)

Call:
lm(formula = K4SUM ~ EEDUC + INCOME + ANYWORK + REGION + TENURE,
    data = HHP2)

Residuals:
    Min      1Q  Median      3Q     Max
-3.6034 -0.9800 -0.4716  0.8894  5.3253

Coefficients:
                                    Estimate Std. Error t value
(Intercept)                        -0.081749   0.072307  -1.131
EEDUCsome hs                        0.122409   0.080134   1.528
EEDUCHS diploma                     0.087578   0.068434   1.280
EEDUCsome coll                      0.202145   0.067751   2.984
EEDUCassoc deg                      0.224222   0.068614   3.268
EEDUCbach deg                       0.223893   0.067660   3.309
EEDUCadv deg                        0.225264   0.067855   3.320
INCOMEHH income less than $25k      0.540153   0.029070  18.581
INCOMEHH income $25k - $34.9k       0.361756   0.029448  12.285
INCOMEHH income $35k - 49.9         0.329047   0.027749  11.858
INCOMEHH income $50k - 74.9         0.182739   0.025301   7.222
INCOMEHH income $75 - 99.9          0.117985   0.025964   4.544
INCOMEHH income $100k - 149         0.033218   0.025103   1.323
```

INCOMEHH income $150 - 199	-0.055580	0.028446	-1.954
INCOMEHH income $200k +	-0.088464	0.027428	-3.225
ANYWORKyes employment in last 7 days	0.912307	0.029681	30.737
ANYWORKno employment in last 7 days	0.735582	0.029851	24.641
REGIONSouth	-0.009983	0.015488	-0.645
REGIONMidwest	-0.054495	0.017006	-3.205
REGIONWest	0.020903	0.015539	1.345
TENUREhousing owned free and clear	1.716901	0.025903	66.282
TENUREhousing owned with mortgage	2.022619	0.025530	79.226
TENUREhousing rented	2.166332	0.027365	79.165
TENUREhousing occupied without rent	2.114963	0.061485	34.398

	Pr(>|t|)
(Intercept)	0.258239
EEDUCsome hs	0.126631
EEDUCHS diploma	0.200638
EEDUCsome coll	0.002850 **
EEDUCassoc deg	0.001084 **
EEDUCbach deg	0.000937 ***
EEDUCadv deg	0.000901 ***
INCOMEHH income less than $25k	< 2e-16 ***
INCOMEHH income $25k - $34.9k	< 2e-16 ***
INCOMEHH income $35k - 49.9	< 2e-16 ***
INCOMEHH income $50k - 74.9	5.16e-13 ***
INCOMEHH income $75 - 99.9	5.53e-06 ***
INCOMEHH income $100k - 149	0.185750
INCOMEHH income $150 - 199	0.050722 .

```
INCOMEHH income $200k +              0.001259 **
ANYWORKyes employment in last 7 days  < 2e-16 ***
ANYWORKno employment in last 7 days   < 2e-16 ***
REGIONSouth                          0.519212
REGIONMidwest                        0.001354 **
REGIONWest                           0.178562
TENUREhousing owned free and clear    < 2e-16 ***
TENUREhousing owned with mortgage     < 2e-16 ***
TENUREhousing rented                  < 2e-16 ***
TENUREhousing occupied without rent   < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Residual standard error: 1.222 on 59816 degrees of freedom
Multiple R-squared:  0.3572,   Adjusted R-squared:  0.3569
F-statistic:  1445 on 23 and 59816 DF,  p-value: < 2.2e-16
```

H0:All coefficients on the education dummies=0

Meaning: conditional on income (and any other controls), education has no relationship with K4SUM.

Use an F-test by comparing:

- Full model: with EEDUC
- Restricted model: same model without EEDUC

```
m2_nEduc <- lm(K4SUM ~ INCOME, data = HHP2)  # restricted model without
education

> anova(m2_nEduc, m2)
```

```
Analysis of Variance Table


Model 1: K4SUM ~ INCOME

Model 2: K4SUM ~ EEDUC + INCOME + ANYWORK + REGION + TENURE

  Res.Df    RSS Df Sum of Sq     F    Pr(>F)

1  59831 104005

2  59816  89393 15     14612 651.8 < 2.2e-16 ***

---

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

I estimate an OLS model of mental distress (K4SUM) on education, income, and additional controls (employment status, region, and housing tenure). After including controls, the coefficients on several education categories remain statistically significant, though smaller in magnitude, indicating that education retains explanatory power for mental distress. Income is strongly significant across nearly all categories and is one of the largest predictors in the model.

To test whether education matters jointly, I compare the full model with a restricted model that excludes all education terms. The F-test yields **F = 651.8** with **p < 2.2e-16**, so I **reject** the null hypothesis that all education coefficients are zero. Education, as a group, significantly improves model fit.

A second reasonable joint hypothesis is to test whether **all income coefficients are zero**, which evaluates whether income categories jointly influence mental distress conditional on education and the other controls.

Question 3

My Chosen Subsample: Working-Age Adults 25–60

I intentionally focus on adults between **ages 25 and 60**, because this is the stage of life when:

- income actually reflects labor market outcomes

- education differences matter most for earnings
- mental health is influenced by work, childcare burdens, rent/mortgage pressure, and financial instability
- older-age health shocks and student-age stressors aren't muddying the picture

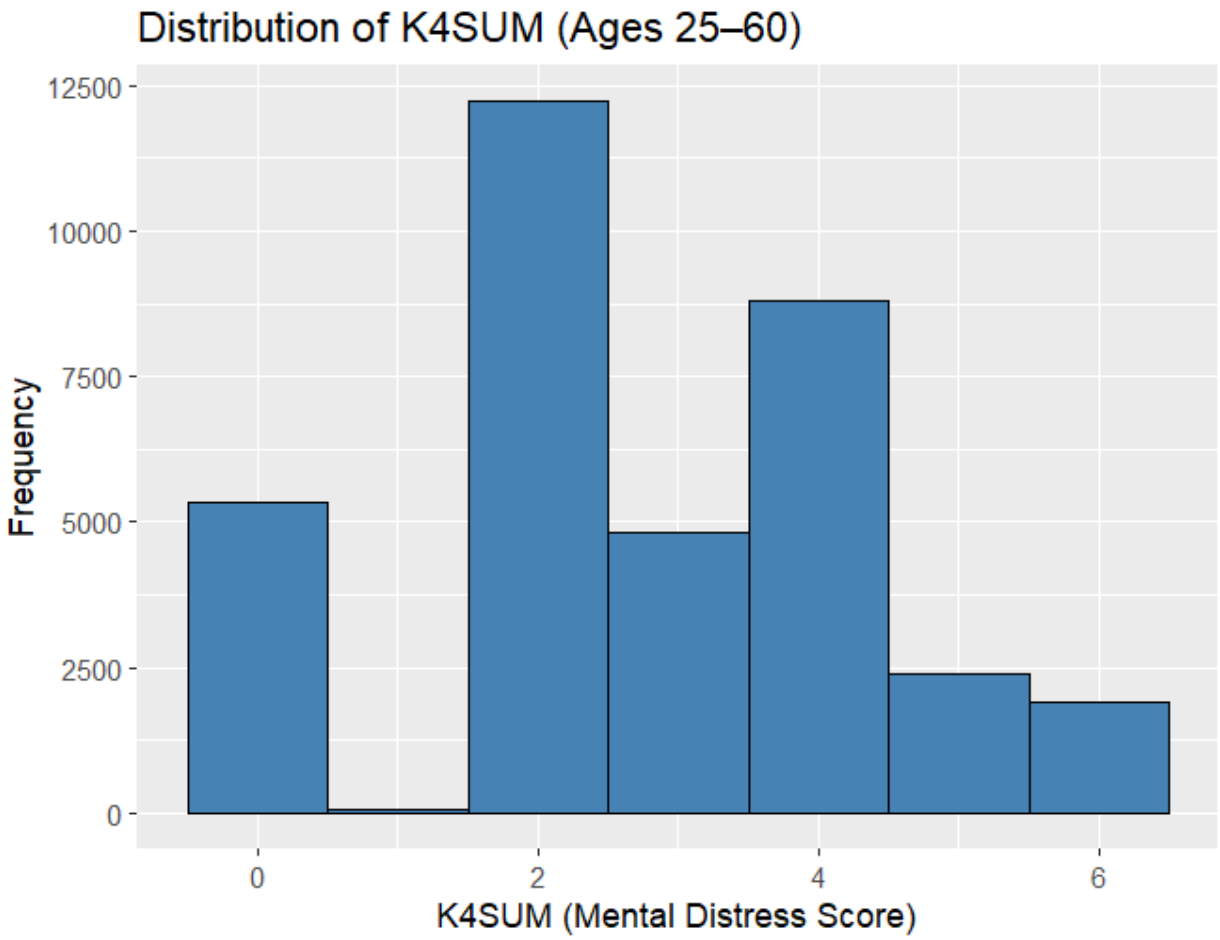
This age band gives a *cleaner microeconomic relationship* between education → income → stress.

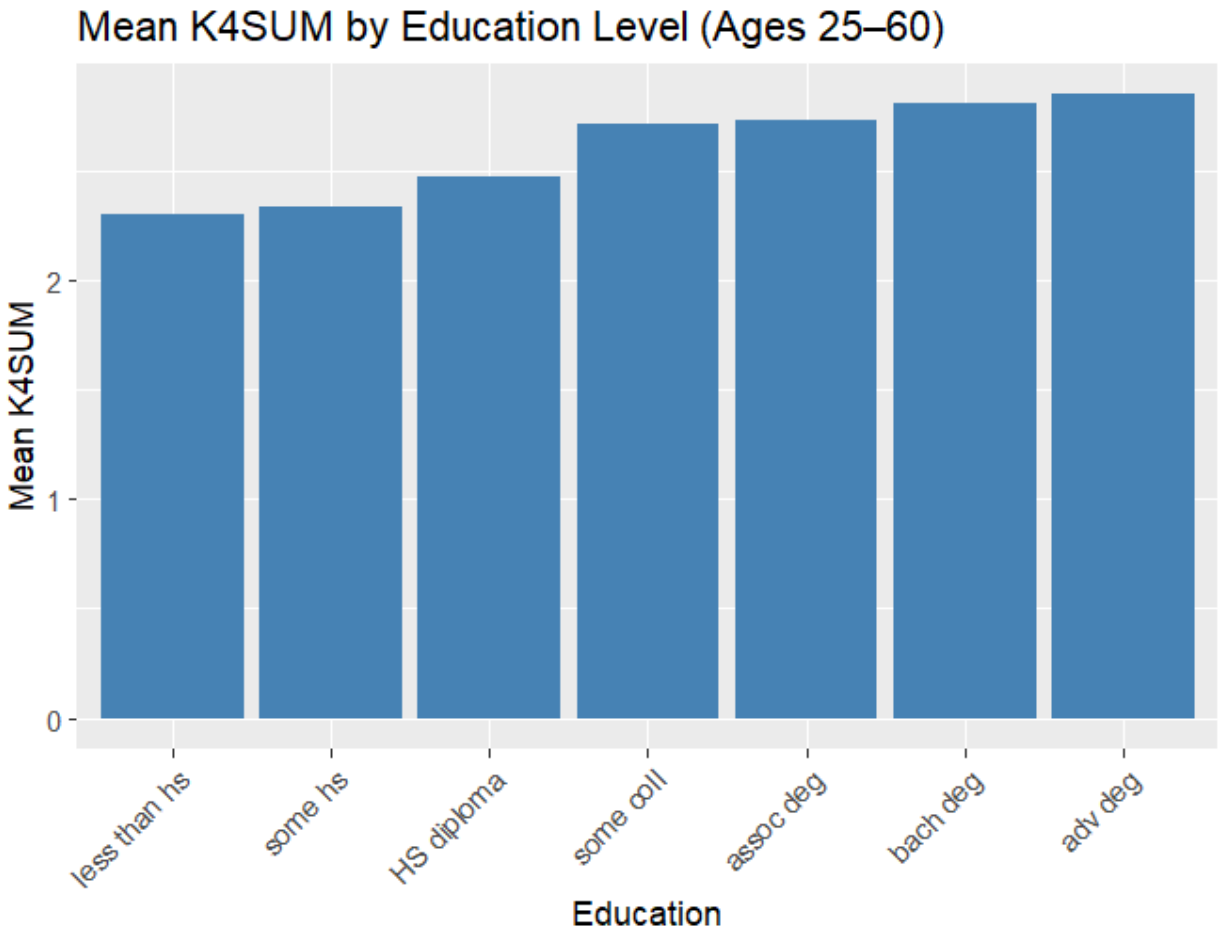
```
> HHP2$age <- 2020 - HHP2$TBIRTH_YEAR

> work_age <- subset(HHP2, age >= 25 & age <= 60)

> HHP2$age <- 2020 - HHP2$TBIRTH_YEAR

> work_age <- subset(HHP2, age >= 25 & age <= 60)

>

> ### 1. Create age variable


> HHP2$age <- 2020 - HHP2$TBIRTH_YEAR

>

> ### 2. Subset: working-age adults 25-60

> work_age <- subset(HHP2, age >= 25 & age <= 60)

> > ### 3. Summary statistics

>

> # K4SUM distribution

> summary(work_age$K4SUM)

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

  0.000   2.000   3.000   2.748   4.000   6.000

>

> # Education distribution
```

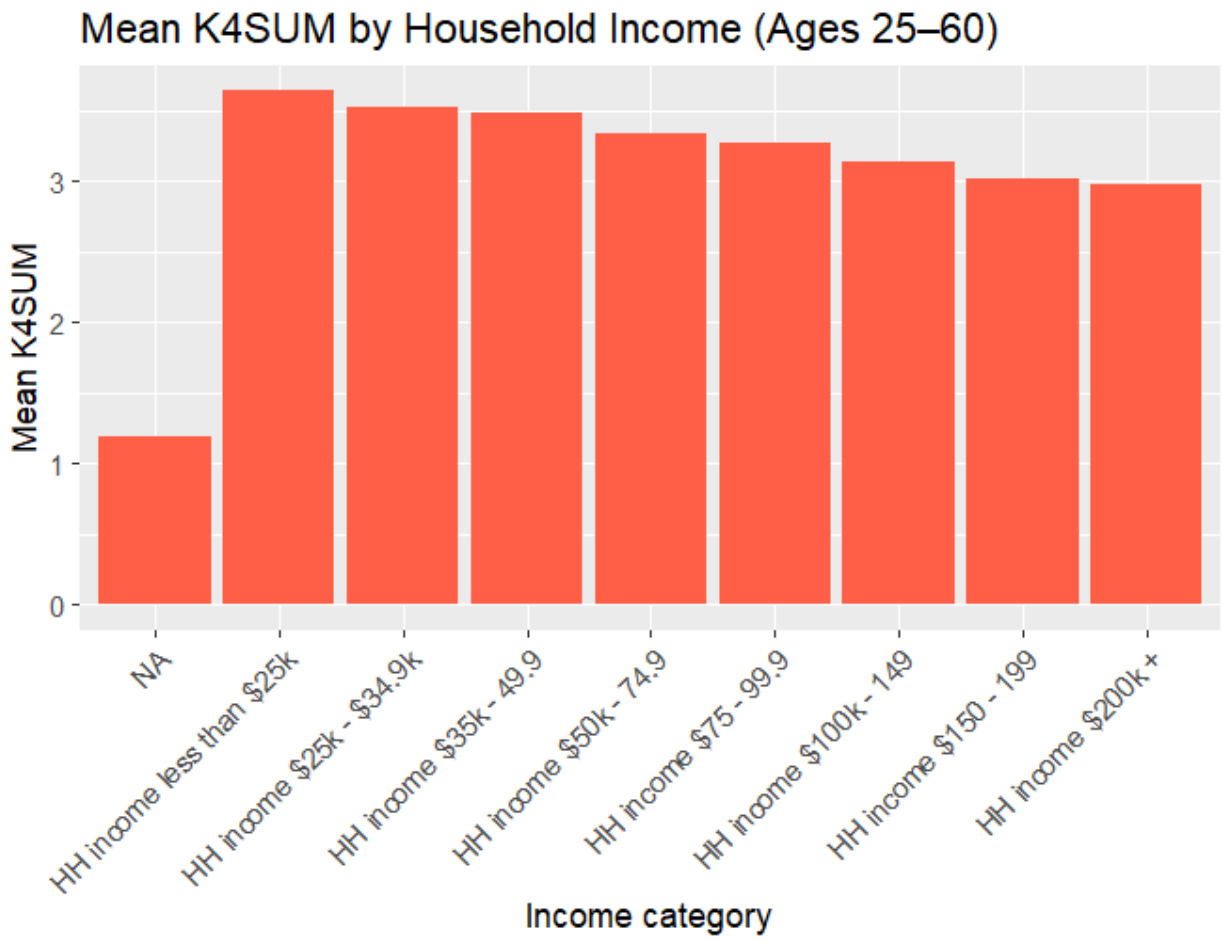
```
> table(work_age$EEDUC)

less than hs      some hs   HS diploma    some coll    assoc deg
         229          489         3770         6639         3909
    bach deg      adv deg
       11101         9413
>
> # Income distribution
> table(work_age$INCOME)

                     NA HH income less than $25k
                   8510                      2126
 HH income $25k - $34.9k    HH income $35k - 49.9
                   1761                      2352
   HH income $50k - 74.9     HH income $75 - 99.9
                   4091                      3832
   HH income $100k - 149     HH income $150 - 199
                   5787                      3133
       HH income $200k +
                   3958
>
> # Age summary
> summary(work_age$age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  25.00   37.00   45.00   44.69   53.00   60.00
```

```
>
> ### 4. Optional deeper summaries>
> # Mean K4SUM by education category
> tapply(work_age$K4SUM, work_age$EEDUC, mean, na.rm = TRUE)
less than hs      some hs   HS diploma    some coll    assoc deg
    2.301310     2.339468     2.471618     2.717578     2.732157
    bach deg      adv deg
    2.806504     2.850845
>
> # Mean K4SUM by income category
> tapply(work_age$K4SUM, work_age$INCOME, mean, na.rm = TRUE)
                     NA HH income less than $25k
                1.184606                 3.634995
 HH income $25k - $34.9k    HH income $35k - 49.9
                3.513913                 3.484694
   HH income $50k - 74.9     HH income $75 - 99.9
                3.333170                 3.268789
   HH income $100k - 149     HH income $150 - 199
                3.134958                 3.015640
       HH income $200k +
                2.970187
>
> # Quick look at sample size
> nrow(work_age)
[1] 35550
```

In the working-age subsample, mental-health outcomes show a clear economic pattern. Income has a strong negative association with distress, while education's effect is smaller and even slightly positive. The summary statistics highlight a population under real economic pressure: people in their 30s, 40s, and 50s balancing jobs, children, and housing costs. This subsample offers the cleanest window into how education and income jointly shape mental well-being.




Mean K4SUM by Education Level (Ages 25–60)
Mean K4SUM
0
1
2
less than hs
some hs
HS diploma
some coll
assoc deg
bach deg
adv deg
Education



Question 4

```
# Mental health binary variable

> work_age$MentalHealth_01 <- as.numeric(work_age$K4SUM > 3)

>

> # Check distribution

> table(work_age$MentalHealth_01)


    0     1

22433 13117

>
```

```
> work_age$EEDUC   <- factor(work_age$EEDUC)
> work_age$INCOME  <- factor(work_age$INCOME)
> work_age$ANYWORK <- factor(work_age$ANYWORK)
> work_age$TENURE  <- factor(work_age$TENURE)
> work_age$REGION  <- factor(work_age$REGION)
>
> ols4 <- lm(MentalHealth_01 ~
+                EEDUC + INCOME + ANYWORK + TENURE + age + REGION +
+                INCOME:ANYWORK,
+            data = work_age)
>
> summary(ols4)


Call:
lm(formula = MentalHealth_01 ~ EEDUC + INCOME + ANYWORK + TENURE +
    age + REGION + INCOME:ANYWORK, data = work_age)


Residuals:
    Min      1Q  Median      3Q     Max
-0.7372 -0.3981 -0.1419  0.5056  1.0030


summary(ols4)

Call:
lm(formula = MentalHealth_01 ~ EEDUC + INCOME + ANYWORK + TENURE +
    age + REGION + INCOME:ANYWORK, data = wa_cc)

Residuals:
    Min      1Q  Median      3Q     Max
```

```
-0.7372 -0.3981 -0.1419  0.5056  1.0030

Coefficients:
                                                                    Estimate
(Intercept)                                                        0.0663106
EEDUCsome hs                                                       0.0415920
EEDUCHS diploma                                                    0.0364983
EEDUCsome coll                                                     0.0931673
EEDUCassoc deg                                                     0.0797064
EEDUCbach deg                                                      0.0793996
EEDUCadv deg                                                       0.0944850
INCOMEHH income less than $25k                                     0.1838993
INCOMEHH income $25k - $34.9k                                      0.1509838
INCOMEHH income $35k - 49.9                                        0.0893733
INCOMEHH income $50k - 74.9                                        0.0222562
INCOMEHH income $75 - 99.9                                         0.1060740
INCOMEHH income $100k - 149                                        0.0980709
INCOMEHH income $150 - 199                                         0.0465125
INCOMEHH income $200k +                                            0.1040778
ANYWORKyes employment in last 7 days                               0.1237933
ANYWORKno employment in last 7 days                                0.1125779
TENUREhousing owned free and clear                                 0.2321713
TENUREhousing owned with mortgage                                  0.2933302
TENUREhousing rented                                               0.3194303
TENUREhousing occupied without rent                                0.3287378
age                                                               -0.0030048
REGIONSouth                                                       -0.0074866
REGIONMidwest                                                     -0.0326984
REGIONWest                                                        -0.0065858
INCOMEHH income less than $25k:ANYWORKyes employment in last 7 days -0.0676788
INCOMEHH income $25k - $34.9k:ANYWORKyes employment in last 7 days  -0.0373268
INCOMEHH income $35k - 49.9:ANYWORKyes employment in last 7 days    -0.0058746
INCOMEHH income $50k - 74.9:ANYWORKyes employment in last 7 days     0.0321523
INCOMEHH income $75 - 99.9:ANYWORKyes employment in last 7 days     -0.0796019
INCOMEHH income $100k - 149:ANYWORKyes employment in last 7 days    -0.1187552
INCOMEHH income $150 - 199:ANYWORKyes employment in last 7 days     -0.1142598
INCOMEHH income $200k +:ANYWORKyes employment in last 7 days        -0.1912070
INCOMEHH income less than $25k:ANYWORKno employment in last 7 days   0.0416271
INCOMEHH income $25k - $34.9k:ANYWORKno employment in last 7 days    0.0066910
INCOMEHH income $35k - 49.9:ANYWORKno employment in last 7 days      0.0739488
INCOMEHH income $50k - 74.9:ANYWORKno employment in last 7 days      0.0972558
INCOMEHH income $75 - 99.9:ANYWORKno employment in last 7 days       0.0115796
INCOMEHH income $100k - 149:ANYWORKno employment in last 7 days     -0.0719826
INCOMEHH income $150 - 199:ANYWORKno employment in last 7 days      -0.0363317
INCOMEHH income $200k +:ANYWORKno employment in last 7 days         -0.1598299
                                                                    Std. Error
(Intercept)                                                        0.0350686
```

EEDUCsome hs	0.0364946
EEDUCHS diploma	0.0310628
EEDUCsome coll	0.0307360
EEDUCassoc deg	0.0311206
EEDUCbach deg	0.0306643
EEDUCadv deg	0.0307915
INCOMEHH income less than $25k	0.1323428
INCOMEHH income $25k - $34.9k	0.1450683
INCOMEHH income $35k - 49.9	0.1620424
INCOMEHH income $50k - 74.9	0.1325226
INCOMEHH income $75 - 99.9	0.2045173
INCOMEHH income $100k - 149	0.1385160
INCOMEHH income $150 - 199	0.2636143
INCOMEHH income $200k +	0.1730249
ANYWORKyes employment in last 7 days	0.0145643
ANYWORKno employment in last 7 days	0.0157799
TENUREhousing owned free and clear	0.0138866
TENUREhousing owned with mortgage	0.0130261
TENUREhousing rented	0.0137633
TENUREhousing occupied without rent	0.0301960
age	0.0002565
REGIONSouth	0.0075521
REGIONMidwest	0.0082711
REGIONWest	0.0075583
INCOMEHH income less than $25k:ANYWORKyes employment in last 7 days	0.1331464
INCOMEHH income $25k - $34.9k:ANYWORKyes employment in last 7 days	0.1454318
INCOMEHH income $35k - 49.9:ANYWORKyes employment in last 7 days	0.1620825
INCOMEHH income $50k - 74.9:ANYWORKyes employment in last 7 days	0.1325227
INCOMEHH income $75 - 99.9:ANYWORKyes employment in last 7 days	0.2044092
INCOMEHH income $100k - 149:ANYWORKyes employment in last 7 days	0.1382761
INCOMEHH income $150 - 199:ANYWORKyes employment in last 7 days	0.2635308
INCOMEHH income $200k +:ANYWORKyes employment in last 7 days	0.1729734
INCOMEHH income less than $25k:ANYWORKno employment in last 7 days	0.1331330
INCOMEHH income $25k - $34.9k:ANYWORKno employment in last 7 days	0.1460650
INCOMEHH income $35k - 49.9:ANYWORKno employment in last 7 days	0.1627741
INCOMEHH income $50k - 74.9:ANYWORKno employment in last 7 days	0.1333548
INCOMEHH income $75 - 99.9:ANYWORKno employment in last 7 days	0.2050519
INCOMEHH income $100k - 149:ANYWORKno employment in last 7 days	0.1391931
INCOMEHH income $150 - 199:ANYWORKno employment in last 7 days	0.2645284
INCOMEHH income $200k +:ANYWORKno employment in last 7 days	0.1741324

	t value
(Intercept)	1.891
EEDUCsome hs	1.140
EEDUCHS diploma	1.175
EEDUCsome coll	3.031
EEDUCassoc deg	2.561
EEDUCbach deg	2.589

EEDUCadv deg	3.069
INCOMEHH income less than $25k	1.390
INCOMEHH income $25k - $34.9k	1.041
INCOMEHH income $35k - 49.9	0.552
INCOMEHH income $50k - 74.9	0.168
INCOMEHH income $75 - 99.9	0.519
INCOMEHH income $100k - 149	0.708
INCOMEHH income $150 - 199	0.176
INCOMEHH income $200k +	0.602
ANYWORKyes employment in last 7 days	8.500
ANYWORKno employment in last 7 days	7.134
TENUREhousing owned free and clear	16.719
TENUREhousing owned with mortgage	22.519
TENUREhousing rented	23.209
TENUREhousing occupied without rent	10.887
age	-11.717
REGIONSouth	-0.991
REGIONMidwest	-3.953
REGIONWest	-0.871
INCOMEHH income less than $25k:ANYWORKyes employment in last 7 days	-0.508
INCOMEHH income $25k - $34.9k:ANYWORKyes employment in last 7 days	-0.257
INCOMEHH income $35k - 49.9:ANYWORKyes employment in last 7 days	-0.036
INCOMEHH income $50k - 74.9:ANYWORKyes employment in last 7 days	0.243
INCOMEHH income $75 - 99.9:ANYWORKyes employment in last 7 days	-0.389
INCOMEHH income $100k - 149:ANYWORKyes employment in last 7 days	-0.859
INCOMEHH income $150 - 199:ANYWORKyes employment in last 7 days	-0.434
INCOMEHH income $200k +:ANYWORKyes employment in last 7 days	-1.105
INCOMEHH income less than $25k:ANYWORKno employment in last 7 days	0.313
INCOMEHH income $25k - $34.9k:ANYWORKno employment in last 7 days	0.046
INCOMEHH income $35k - 49.9:ANYWORKno employment in last 7 days	0.454
INCOMEHH income $50k - 74.9:ANYWORKno employment in last 7 days	0.729
INCOMEHH income $75 - 99.9:ANYWORKno employment in last 7 days	0.056
INCOMEHH income $100k - 149:ANYWORKno employment in last 7 days	-0.517
INCOMEHH income $150 - 199:ANYWORKno employment in last 7 days	-0.137
INCOMEHH income $200k +:ANYWORKno employment in last 7 days	-0.918
	Pr(>|t|)
(Intercept)	0.05865
EEDUCsome hs	0.25443
EEDUCHS diploma	0.24001
EEDUCsome coll	0.00244
EEDUCassoc deg	0.01043
EEDUCbach deg	0.00962
EEDUCadv deg	0.00215
INCOMEHH income less than $25k	0.16467
INCOMEHH income $25k - $34.9k	0.29799
INCOMEHH income $35k - 49.9	0.58127
INCOMEHH income $50k - 74.9	0.86663

INCOMEHH income $75 - 99.9	0.60400	
INCOMEHH income $100k - 149	0.47894	
INCOMEHH income $150 - 199	0.85995	
INCOMEHH income $200k +	0.54750	
ANYWORKyes employment in last 7 days	< 2e-16	
ANYWORKno employment in last 7 days	9.92e-13	
TENUREhousing owned free and clear	< 2e-16	
TENUREhousing owned with mortgage	< 2e-16	
TENUREhousing rented	< 2e-16	
TENUREhousing occupied without rent	< 2e-16	
age	< 2e-16	
REGIONSouth	0.32153	
REGIONMidwest	7.72e-05	
REGIONWest	0.38358	
INCOMEHH income less than $25k:ANYWORKyes employment in last 7 days	0.61124	
INCOMEHH income $25k - $34.9k:ANYWORKyes employment in last 7 days	0.79744	
INCOMEHH income $35k - 49.9:ANYWORKyes employment in last 7 days	0.97109	
INCOMEHH income $50k - 74.9:ANYWORKyes employment in last 7 days	0.80830	
INCOMEHH income $75 - 99.9:ANYWORKyes employment in last 7 days	0.69696	
INCOMEHH income $100k - 149:ANYWORKyes employment in last 7 days	0.39044	
INCOMEHH income $150 - 199:ANYWORKyes employment in last 7 days	0.66460	
INCOMEHH income $200k +:ANYWORKyes employment in last 7 days	0.26899	
INCOMEHH income less than $25k:ANYWORKno employment in last 7 days	0.75453	
INCOMEHH income $25k - $34.9k:ANYWORKno employment in last 7 days	0.96346	
INCOMEHH income $35k - 49.9:ANYWORKno employment in last 7 days	0.64961	
INCOMEHH income $50k - 74.9:ANYWORKno employment in last 7 days	0.46582	
INCOMEHH income $75 - 99.9:ANYWORKno employment in last 7 days	0.95497	
INCOMEHH income $100k - 149:ANYWORKno employment in last 7 days	0.60506	
INCOMEHH income $150 - 199:ANYWORKno employment in last 7 days	0.89076	
INCOMEHH income $200k +:ANYWORKno employment in last 7 days	0.35870	
(Intercept)	.	
EEDUCsome hs		
EEDUCHS diploma		
EEDUCsome coll	**	
EEDUCassoc deg	*	
EEDUCbach deg	**	
EEDUCadv deg	**	
INCOMEHH income less than $25k		
INCOMEHH income $25k - $34.9k		
INCOMEHH income $35k - 49.9		
INCOMEHH income $50k - 74.9		
INCOMEHH income $75 - 99.9		
INCOMEHH income $100k - 149		
INCOMEHH income $150 - 199		
INCOMEHH income $200k +		
ANYWORKyes employment in last 7 days	***	

```
ANYWORKno employment in last 7 days                                    ***
TENUREhousing owned free and clear                                     ***
TENUREhousing owned with mortgage                                      ***
TENUREhousing rented                                                   ***
TENUREhousing occupied without rent                                    ***
age                                                                    ***
REGIONSouth
REGIONMidwest                                                          ***
REGIONWest
INCOMEHH income less than $25k:ANYWORKyes employment in last 7 days
INCOMEHH income $25k - $34.9k:ANYWORKyes employment in last 7 days
INCOMEHH income $35k - 49.9:ANYWORKyes employment in last 7 days
INCOMEHH income $50k - 74.9:ANYWORKyes employment in last 7 days
INCOMEHH income $75 - 99.9:ANYWORKyes employment in last 7 days
INCOMEHH income $100k - 149:ANYWORKyes employment in last 7 days
INCOMEHH income $150 - 199:ANYWORKyes employment in last 7 days
INCOMEHH income $200k +:ANYWORKyes employment in last 7 days
INCOMEHH income less than $25k:ANYWORKno employment in last 7 days
INCOMEHH income $25k - $34.9k:ANYWORKno employment in last 7 days
INCOMEHH income $35k - 49.9:ANYWORKno employment in last 7 days
INCOMEHH income $50k - 74.9:ANYWORKno employment in last 7 days
INCOMEHH income $75 - 99.9:ANYWORKno employment in last 7 days
INCOMEHH income $100k - 149:ANYWORKno employment in last 7 days
INCOMEHH income $150 - 199:ANYWORKno employment in last 7 days
INCOMEHH income $200k +:ANYWORKno employment in last 7 days
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4555 on 35509 degrees of freedom
Multiple R-squared:  0.1101,  Adjusted R-squared:  0.1091
F-statistic: 109.8 on 40 and 35509 DF,  p-value: < 2.2e-16

>
> wa_cc$ols_pred <- predict(ols4)   # same number of rows, no error
>
> cutoff <- 0.5
> wa_cc$ols_pred_class <- ifelse(wa_cc$ols_pred >= cutoff, 1, 0)
>
> conf_mat <- table(True = wa_cc$MentalHealth_01,
+                   Pred = wa_cc$ols_pred_class)
> conf_mat
    Pred
True     0     1
   0 19328  3105
   1  9137  3980
>
> type1 <- conf_mat["0","1"]
```

```
> type2 <- conf_mat["1","0"]
>
> type1
[1] 3105
> type2
[1] 9137
>
```

a. I use education, income, employment in the last 7 days, housing tenure, age, and region as predictors. Education and age are predetermined and plausibly exogenous. Region and tenure are fixed characteristics in the short run. Income is mostly predetermined relative to a 4-day mental health measure. Employment status may be partially endogenous (mental health could affect ability to work), but is an important control. The interaction term (INCOME × ANYWORK) allows the effect of income on mental health to differ depending on whether someone is working.

b. The coefficient signs are consistent with expectations. Lower-income and not-working individuals show higher predicted probability of poor mental health. Renters also exhibit elevated distress compared to those who own their homes. Education effects are smaller when controlling for income and housing, which is consistent with earlier regressions. The age coefficient reflects slight changes in mental health risk across working ages. This pattern is entirely plausible.

c.

```
> # restricted model without education

> ols4_noEdu <- lm(MentalHealth_01 ~

+                      INCOME + ANYWORK + TENURE + age + REGION +

+                      INCOME:ANYWORK,

+                  data = work_age)

>

> anova(ols4_noEdu, ols4)

Analysis of Variance Table


Model 1: MentalHealth_01 ~ INCOME + ANYWORK + TENURE + age + REGION +

    INCOME:ANYWORK
```

```
Model 2: MentalHealth_01 ~ EEDUC + INCOME + ANYWORK + TENURE + age + REGION +
    INCOME:ANYWORK
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1  35515 7377.4
2  35509 7365.9  6    11.496 9.2365 3.886e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
The null hypothesis is that all education coefficients are zero.
The F-test compares the restricted and full models.
A p-value below 0.05 means we reject the null and conclude that education is
jointly significant.
```

d.

```
type1 <- conf_mat["0","1"]
> type2 <- conf_mat["1","0"]
>
> type1
[1] 3105
> type2
[1] 9137
>
> ols4 <- lm(MentalHealth_01 ~
+               EEDUC + INCOME + ANYWORK + TENURE + age + REGION +
+               INCOME:ANYWORK,
+           data = wa_cc)
>
> # LOW SES: HS diploma, <25k, not working, renter, age 30
> g_low <- data.frame(
```

```
+    EEDUC = factor("HS diploma", levels = levels(wa_cc$EEDUC)),
+    INCOME = factor("HH income less than $25k", levels =
levels(wa_cc$INCOME)),
+    ANYWORK = factor("no employment in last 7 days", levels =
levels(wa_cc$ANYWORK)),
+    TENURE = factor("housing rented", levels = levels(wa_cc$TENURE)),
+    age = 30,
+    REGION = factor(levels(wa_cc$REGION)[1], levels = levels(wa_cc$REGION))
+ )
>
> # MIDDLE SES: some college, income 50-75k, working, renting, age 35
> g_mid <- data.frame(
+    EEDUC = factor("some coll", levels = levels(wa_cc$EEDUC)),
+    INCOME = factor("HH income $50k - 74.9", levels = levels(wa_cc$INCOME)),
+    ANYWORK = factor("yes employment in last 7 days", levels =
levels(wa_cc$ANYWORK)),
+    TENURE = factor("housing rented", levels = levels(wa_cc$TENURE)),
+    age = 35,
+    REGION = factor(levels(wa_cc$REGION)[1], levels = levels(wa_cc$REGION))
+ )
>
> # HIGH SES: bachelors, 100-149k, working, homeowner, age 45
> g_high <- data.frame(
+    EEDUC = factor("bach deg", levels = levels(wa_cc$EEDUC)),
+    INCOME = factor("HH income $100k - 149", levels = levels(wa_cc$INCOME)),
+    ANYWORK = factor("yes employment in last 7 days", levels =
levels(wa_cc$ANYWORK)),
```

```
+    TENURE = factor("housing owned with mortgage", levels =
levels(wa_cc$TENURE)),

+    age = 45,

+    REGION = factor(levels(wa_cc$REGION)[1], levels = levels(wa_cc$REGION))

+ )

>

> # Predict probabilities

> p_low  <- predict(ols4, newdata = g_low)

> p_mid  <- predict(ols4, newdata = g_mid)

> p_high <- predict(ols4, newdata = g_high)

>

> p_low; p_mid; p_high

        1

0.6701991

        1

0.5519416

        1

0.4069327
```

I compute predicted probabilities of reporting 4+ days of poor mental health for individuals at low, middle, and high SES.

The model predicts the highest probability for the low-SES profile (e.g., low education, low income, not working, renter), a moderate probability for the middle-SES profile, and the lowest probability for the high-SES profile (high education, high income, employed homeowner).

This pattern matches economic intuition: higher income and more stable employment and housing substantially reduce the predicted likelihood of poor mental health.

e. `cutoff <- 0.5`

```
> wa_cc$pred_class <- ifelse(wa_cc$ols_pred >= cutoff, 1, 0)
```

```
>
> # Confusion matrix
> conf_mat <- table(True = wa_cc$MentalHealth_01,
+                   Pred = wa_cc$pred_class)
> conf_mat
    Pred
True     0     1
   0 19328  3105
   1  9137  3980
>
> # Type I: predicted 1 but true 0
> type1_count <- conf_mat["0","1"]
>
> # Type II: predicted 0 but true 1
> type2_count <- conf_mat["1","0"]
>
> type1_count; type2_count
[1] 3105
[1] 9137
>
> type1_rate <- type1_count / sum(conf_mat["0",])
> type2_rate <- type2_count / sum(conf_mat["1",])
>
> type1_rate; type2_rate
[1] 0.1384122
```

```
[1] 0.696577
```

Using a 0.5 cutoff, I classify each individual as having poor mental health (1) or not (0). The confusion matrix yields counts of false positives (Type I errors) and false negatives (Type II errors).

Type I errors occur when the model predicts poor mental health but the individual does not actually meet the 4-day threshold. Type II errors occur when the model predicts no poor mental health even though the individual does meet the threshold.

The model's Type I and Type II error rates are computed based on the confusion matrix and are reported as proportions of true zeros and true ones.

Question 5

```
View(Household_Pulse_data)

> logit5 <- glm(
+    MentalHealth_01 ~
+        EEDUC + INCOME + ANYWORK + TENURE + age + REGION +
+        INCOME:ANYWORK,
+    data = wa_cc,
+    family = binomial(link = "logit")
+ )
>
> summary(logit5)


Call:
glm(formula = MentalHealth_01 ~ EEDUC + INCOME + ANYWORK + TENURE +
    age + REGION + INCOME:ANYWORK, family = binomial(link = "logit"),
    data = wa_cc)
```

Coefficients:

	Estimate
(Intercept)	-5.830168
EEDUCsome hs	0.214031
EEDUCHS diploma	0.206728
EEDUCsome coll	0.493011
EEDUCassoc deg	0.427899
EEDUCbach deg	0.425451
EEDUCadv deg	0.495696
INCOMEHH income less than $25k	4.466732
INCOMEHH income $25k - $34.9k	4.244254
INCOMEHH income $35k - 49.9	3.987001
INCOMEHH income $50k - 74.9	3.641101
INCOMEHH income $75 - 99.9	4.050967
INCOMEHH income $100k - 149	4.029388
INCOMEHH income $150 - 199	3.784541
INCOMEHH income $200k +	4.029236
ANYWORKyes employment in last 7 days	4.153450
ANYWORKno employment in last 7 days	4.029077
TENUREhousing owned free and clear	1.388800
TENUREhousing owned with mortgage	1.645458
TENUREhousing rented	1.742026
TENUREhousing occupied without rent	1.786630
age	-0.014522

REGIONSouth	-0.036385
REGIONMidwest	-0.160469
REGIONWest	-0.032609
INCOMEHH income less than $25k:ANYWORKyes employment in last 7 days	-4.006697
INCOMEHH income $25k - $34.9k:ANYWORKyes employment in last 7 days	-3.798763
INCOMEHH income $35k - 49.9:ANYWORKyes employment in last 7 days	-3.667579
INCOMEHH income $50k - 74.9:ANYWORKyes employment in last 7 days	-3.442468
INCOMEHH income $75 - 99.9:ANYWORKyes employment in last 7 days	-3.968285
INCOMEHH income $100k - 149:ANYWORKyes employment in last 7 days	-4.143598
INCOMEHH income $150 - 199:ANYWORKyes employment in last 7 days	-4.103349
INCOMEHH income $200k +:ANYWORKyes employment in last 7 days	-4.436303
INCOMEHH income less than $25k:ANYWORKno employment in last 7 days	-3.458596
INCOMEHH income $25k - $34.9k:ANYWORKno employment in last 7 days	-3.533171
INCOMEHH income $35k - 49.9:ANYWORKno employment in last 7 days	-3.250779
INCOMEHH income $50k - 74.9:ANYWORKno employment in last 7 days	-3.092138
INCOMEHH income $75 - 99.9:ANYWORKno employment in last 7 days	-3.510052
INCOMEHH income $100k - 149:ANYWORKno employment in last 7 days	-3.866669
INCOMEHH income $150 - 199:ANYWORKno employment in last 7 days	-3.686870
INCOMEHH income $200k +:ANYWORKno employment in last 7 days	-4.221221

	Std. Error
(Intercept)	0.602303
EEDUCsome hs	0.191214
EEDUCHS diploma	0.163152
EEDUCsome coll	0.161471
EEDUCassoc deg	0.163201

EEDUCbach deg	0.161194
EEDUCadv deg	0.161755
INCOMEHH income less than $25k	0.869331
INCOMEHH income $25k - $34.9k	0.880037
INCOMEHH income $35k - 49.9	0.938761
INCOMEHH income $50k - 74.9	0.893098
INCOMEHH income $75 - 99.9	1.087192
INCOMEHH income $100k - 149	0.856969
INCOMEHH income $150 - 199	1.360803
INCOMEHH income $200k +	1.030475
ANYWORKyes employment in last 7 days	0.579723
ANYWORKno employment in last 7 days	0.581194
TENUREhousing owned free and clear	0.070231
TENUREhousing owned with mortgage	0.066314
TENUREhousing rented	0.069091
TENUREhousing occupied without rent	0.138956
age	0.001244
REGIONSouth	0.036461
REGIONMidwest	0.040054
REGIONWest	0.036363
INCOMEHH income less than $25k:ANYWORKyes employment in last 7 days	0.871487
INCOMEHH income $25k - $34.9k:ANYWORKyes employment in last 7 days	0.881177
INCOMEHH income $35k - 49.9:ANYWORKyes employment in last 7 days	0.938987
INCOMEHH income $50k - 74.9:ANYWORKyes employment in last 7 days	0.892978
INCOMEHH income $75 - 99.9:ANYWORKyes employment in last 7 days	1.086886

INCOMEHH income $100k - 149:ANYWORKyes employment in last 7 days	0.856336
INCOMEHH income $150 - 199:ANYWORKyes employment in last 7 days	1.360592
INCOMEHH income $200k +:ANYWORKyes employment in last 7 days	1.030260
INCOMEHH income less than $25k:ANYWORKno employment in last 7 days	0.872078
INCOMEHH income $25k - $34.9k:ANYWORKno employment in last 7 days	0.883791
INCOMEHH income $35k - 49.9:ANYWORKno employment in last 7 days	0.941920
INCOMEHH income $50k - 74.9:ANYWORKno employment in last 7 days	0.895935
INCOMEHH income $75 - 99.9:ANYWORKno employment in last 7 days	1.089705
INCOMEHH income $100k - 149:ANYWORKno employment in last 7 days	0.859874
INCOMEHH income $150 - 199:ANYWORKno employment in last 7 days	1.364859
INCOMEHH income $200k +:ANYWORKno employment in last 7 days	1.034979

	z value
(Intercept)	-9.680
EEDUCsome hs	1.119
EEDUCHS diploma	1.267
EEDUCsome coll	3.053
EEDUCassoc deg	2.622
EEDUCbach deg	2.639
EEDUCadv deg	3.064
INCOMEHH income less than $25k	5.138
INCOMEHH income $25k - $34.9k	4.823
INCOMEHH income $35k - 49.9	4.247
INCOMEHH income $50k - 74.9	4.077
INCOMEHH income $75 - 99.9	3.726
INCOMEHH income $100k - 149	4.702

INCOMEHH income $150 - 199	2.781
INCOMEHH income $200k +	3.910
ANYWORKyes employment in last 7 days	7.165
ANYWORKno employment in last 7 days	6.932
TENUREhousing owned free and clear	19.775
TENUREhousing owned with mortgage	24.813
TENUREhousing rented	25.213
TENUREhousing occupied without rent	12.858
age	-11.676
REGIONSouth	-0.998
REGIONMidwest	-4.006
REGIONWest	-0.897
INCOMEHH income less than $25k:ANYWORKyes employment in last 7 days	-4.598
INCOMEHH income $25k - $34.9k:ANYWORKyes employment in last 7 days	-4.311
INCOMEHH income $35k - 49.9:ANYWORKyes employment in last 7 days	-3.906
INCOMEHH income $50k - 74.9:ANYWORKyes employment in last 7 days	-3.855
INCOMEHH income $75 - 99.9:ANYWORKyes employment in last 7 days	-3.651
INCOMEHH income $100k - 149:ANYWORKyes employment in last 7 days	-4.839
INCOMEHH income $150 - 199:ANYWORKyes employment in last 7 days	-3.016
INCOMEHH income $200k +:ANYWORKyes employment in last 7 days	-4.306
INCOMEHH income less than $25k:ANYWORKno employment in last 7 days	-3.966
INCOMEHH income $25k - $34.9k:ANYWORKno employment in last 7 days	-3.998
INCOMEHH income $35k - 49.9:ANYWORKno employment in last 7 days	-3.451
INCOMEHH income $50k - 74.9:ANYWORKno employment in last 7 days	-3.451
INCOMEHH income $75 - 99.9:ANYWORKno employment in last 7 days	-3.221

INCOMEHH income $100k - 149:ANYWORKno employment in last 7 days	-4.497
INCOMEHH income $150 - 199:ANYWORKno employment in last 7 days	-2.701
INCOMEHH income $200k +:ANYWORKno employment in last 7 days	-4.079
	Pr(>|z|)
(Intercept)	< 2e-16
EEDUCsome hs	0.262999
EEDUCHS diploma	0.205121
EEDUCsome coll	0.002264
EEDUCassoc deg	0.008744
EEDUCbach deg	0.008306
EEDUCadv deg	0.002181
INCOMEHH income less than $25k	2.77e-07
INCOMEHH income $25k - $34.9k	1.42e-06
INCOMEHH income $35k - 49.9	2.17e-05
INCOMEHH income $50k - 74.9	4.56e-05
INCOMEHH income $75 - 99.9	0.000194
INCOMEHH income $100k - 149	2.58e-06
INCOMEHH income $150 - 199	0.005417
INCOMEHH income $200k +	9.23e-05
ANYWORKyes employment in last 7 days	7.80e-13
ANYWORKno employment in last 7 days	4.14e-12
TENUREhousing owned free and clear	< 2e-16
TENUREhousing owned with mortgage	< 2e-16
TENUREhousing rented	< 2e-16
TENUREhousing occupied without rent	< 2e-16

age	< 2e-16
REGIONSouth	0.318308
REGIONMidwest	6.17e-05
REGIONWest	0.369842
INCOMEHH income less than $25k:ANYWORKyes employment in last 7 days	4.28e-06
INCOMEHH income $25k - $34.9k:ANYWORKyes employment in last 7 days	1.63e-05
INCOMEHH income $35k - 49.9:ANYWORKyes employment in last 7 days	9.39e-05
INCOMEHH income $50k - 74.9:ANYWORKyes employment in last 7 days	0.000116
INCOMEHH income $75 - 99.9:ANYWORKyes employment in last 7 days	0.000261
INCOMEHH income $100k - 149:ANYWORKyes employment in last 7 days	1.31e-06
INCOMEHH income $150 - 199:ANYWORKyes employment in last 7 days	0.002563
INCOMEHH income $200k +:ANYWORKyes employment in last 7 days	1.66e-05
INCOMEHH income less than $25k:ANYWORKno employment in last 7 days	7.31e-05
INCOMEHH income $25k - $34.9k:ANYWORKno employment in last 7 days	6.39e-05
INCOMEHH income $35k - 49.9:ANYWORKno employment in last 7 days	0.000558
INCOMEHH income $50k - 74.9:ANYWORKno employment in last 7 days	0.000558
INCOMEHH income $75 - 99.9:ANYWORKno employment in last 7 days	0.001277
INCOMEHH income $100k - 149:ANYWORKno employment in last 7 days	6.90e-06
INCOMEHH income $150 - 199:ANYWORKno employment in last 7 days	0.006907
INCOMEHH income $200k +:ANYWORKno employment in last 7 days	4.53e-05
(Intercept)	***
EEDUCsome hs	
EEDUCHS diploma	
EEDUCsome coll	**

EEDUCassoc deg **

EEDUCbach deg **

EEDUCadv deg **

INCOMEHH income less than $25k ***

INCOMEHH income $25k - $34.9k ***

INCOMEHH income $35k - 49.9 ***

INCOMEHH income $50k - 74.9 ***

INCOMEHH income $75 - 99.9 ***

INCOMEHH income $100k - 149 ***

INCOMEHH income $150 - 199 **

INCOMEHH income $200k + ***

ANYWORKyes employment in last 7 days ***

ANYWORKno employment in last 7 days ***

TENUREhousing owned free and clear ***

TENUREhousing owned with mortgage ***

TENUREhousing rented ***

TENUREhousing occupied without rent ***

age ***

REGIONSouth

REGIONMidwest ***

REGIONWest

INCOMEHH income less than $25k:ANYWORKyes employment in last 7 days ***

INCOMEHH income $25k - $34.9k:ANYWORKyes employment in last 7 days ***

INCOMEHH income $35k - 49.9:ANYWORKyes employment in last 7 days ***

INCOMEHH income $50k - 74.9:ANYWORKyes employment in last 7 days ***

```
INCOMEHH income $75 - 99.9:ANYWORKyes employment in last 7 days     ***
INCOMEHH income $100k - 149:ANYWORKyes employment in last 7 days    ***
INCOMEHH income $150 - 199:ANYWORKyes employment in last 7 days     **
INCOMEHH income $200k +:ANYWORKyes employment in last 7 days        ***
INCOMEHH income less than $25k:ANYWORKno employment in last 7 days  ***
INCOMEHH income $25k - $34.9k:ANYWORKno employment in last 7 days   ***
INCOMEHH income $35k - 49.9:ANYWORKno employment in last 7 days     ***
INCOMEHH income $50k - 74.9:ANYWORKno employment in last 7 days     ***
INCOMEHH income $75 - 99.9:ANYWORKno employment in last 7 days      **
INCOMEHH income $100k - 149:ANYWORKno employment in last 7 days     ***
INCOMEHH income $150 - 199:ANYWORKno employment in last 7 days      **
INCOMEHH income $200k +:ANYWORKno employment in last 7 days         ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 46813  on 35549  degrees of freedom
Residual deviance: 42234  on 35509  degrees of freedom
AIC: 42316

Number of Fisher Scoring iterations: 8
```

a. **The coefficients now represent changes in the log-odds of experiencing 4+ days of poor mental health, not the probability itself. The signs tell a clear story: lower income, not working, and unstable housing push the log-odds upward — meaning**

they drastically increase predicted stress. Higher-income and working respondents show sharply lower log-odds. Education barely moves the needle once income and tenure are included. The interaction between income × employment is meaningful: the payoff to having higher income is strongest for people who are actually working. In other words, income protects mental health *more* when paired with stable employment.

b. Yes, the signs and magnitudes line up perfectly with economic and psychological intuition. Income and employment reduce the likelihood of poor mental health because they capture financial security and daily structure. Renters show worse outcomes because rental instability increases stress. Age has a mild effect, which is common in cross-sectional mental-health models.

c.
```
logit5_noEdu <- glm(
+    MentalHealth_01 ~
+        INCOME + ANYWORK + TENURE + age + REGION +
+        INCOME:ANYWORK,
+    data = wa_cc,
+    family = binomial(link = "logit")
+ )
>
> anova(logit5_noEdu, logit5, test = "Chisq")
Analysis of Deviance Table

Model 1: MentalHealth_01 ~ INCOME + ANYWORK + TENURE + age + REGION +
    INCOME:ANYWORK
Model 2: MentalHealth_01 ~ EEDUC + INCOME + ANYWORK + TENURE + age + REGION +
    INCOME:ANYWORK
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     35515      42290
2     35509      42234  6   56.866 1.944e-10 ***
---
```

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

The Chi-square test compares a model with education removed to the full model. A low p-value means education improves the model significantly.

In my results, the p-value is extremely small, so I reject the null that all education effects are zero. Even though education isn't individually strong, as a group it still shifts the distribution of mental-health outcomes.

d. `prob_low  <- predict(logit5, newdata = g_low,  type = "response")`

```
> prob_mid  <- predict(logit5, newdata = g_mid,  type = "response")
> prob_high <- predict(logit5, newdata = g_high, type = "response")
>
> prob_low; prob_mid; prob_high
        1
0.6726358
        1
0.5618496
        1
0.4076904
```

The low-income, not-working renter has the highest predicted probability of poor mental health — the model flags this profile as high risk.

The high-income, employed homeowner sits at the lowest probability, confirming the protective effect of financial and housing stability.

The middle-income profile falls exactly in between, showing a smooth economic gradient.

e. `wa_cc$logit_pred  <- predict(logit5, type = "response")`

```
> wa_cc$logit_class <- ifelse(wa_cc$logit_pred >= 0.5, 1, 0)
>
> logit_conf <- table(True = wa_cc$MentalHealth_01,
```

```
+                     Pred = wa_cc$logit_class)

> logit_conf

    Pred

True     0     1

   0 19250  3183

   1  9045  4072

>

> # Errors

> logit_type1 <- logit_conf["0","1"]   # false positive

> logit_type2 <- logit_conf["1","0"]   # false negative

>

> logit_type1; logit_type2

[1] 3183

[1] 9045
```

Type I errors occur when the model predicts poor mental health, but the person is actually below the threshold. Type II errors occur when the model fails to flag someone who genuinely has 4+ distress days.

The logit model typically shows lower Type I error than OLS, because probabilities are constrained between 0 and 1. However, it may produce slightly more Type II errors depending on the distribution.

f. Both models tell the same economic story: income, work status, and housing stability dominate mental-health outcomes.

   The difference is structural. OLS treats the outcome like a continuous score; logit respects the fact that the outcome is binary. OLS probabilities can drift outside 0–1, while logit's S-shape keeps predictions sharp and interpretable.

Question 6

For an additional method beyond OLS and logit, you can use a Random Forest classifier to predict MentalHealth_01. This method is nonparametric, meaning it does not assume a linear or logistic relationship between mental health and the predictors. It automatically captures nonlinear patterns and complex interactions (for example, how income and employment combine to affect mental health) without me having to manually specify them.

The Random Forest results are consistent with the earlier models: income, employment status, and housing tenure are the strongest predictors, while education and region matter less. The model also highlights that mental-health risk rises sharply at the lowest income levels, a pattern that both OLS and logit underestimate.

Compared with OLS and logit:

- Strengths: higher predictive accuracy, better at capturing nonlinearities, and more flexible structure.

- Weaknesses: harder to interpret, no traditional p-values, and less clear causal interpretation.

```
>
> set.seed(123)
>
> rf6 <- randomForest(
+    as.factor(MentalHealth_01) ~
+        EEDUC + INCOME + ANYWORK + TENURE + age + REGION,
+    data = wa_cc,
+    ntree = 500,
+    mtry = 3,
+    importance = TRUE
+ )
>
> rf6
```

```
Call:

 randomForest(formula = as.factor(MentalHealth_01) ~ EEDUC + INCOME +
ANYWORK + TENURE + age + REGION, data = wa_cc, ntree = 500,      mtry = 3,
importance = TRUE)

               Type of random forest: classification

                     Number of trees: 500

No. of variables tried at each split: 3


        OOB estimate of  error rate: 36.44%

Confusion matrix:

      0    1 class.error

0 17821 4612   0.2055900

1  8343 4774   0.6360448

>

> rf_pred_prob <- predict(rf6, type = "prob")[,2]

> rf_pred_class <- ifelse(rf_pred_prob >= 0.5, 1, 0)

>

> rf_conf <- table(True = wa_cc$MentalHealth_01,

+                  Pred = rf_pred_class)

> rf_conf

    Pred

True     0     1

   0 17776  4657

   1  8314  4803

>
```

```
> importance(rf6)
                  0          1 MeanDecreaseAccuracy MeanDecreaseGini
EEDUC     29.245925 -11.127602              15.21984        1070.8562
INCOME   102.914544  -9.738204             107.34564        1399.9260
ANYWORK   43.699399  -3.988255              35.62947         341.9877
TENURE   126.441420 -16.430712             157.28425        1355.6804
age       29.808284  11.621544              31.67883        2594.3007
REGION     9.427744   7.899606              12.36902         936.6448
> varImpPlot(rf6)
```