

```

library(tidyverse) library(modelsummary) library(readr) library(ggplot2)

data <- read_csv("GSSdata2018.csv") View(data)

head(data) dim(data)

summary(data)

data <- data %>% mutate(income = 500(RINCOME == 1) + 1000(RINCOME == 2) + 3000(RINCOME
== 3) + 4000(RINCOME == 4) + 5000(RINCOME == 5) + 6000(RINCOME == 6) + 7000(RINCOME
== 7) + 8000(RINCOME == 8) + 10000(RINCOME == 9) + 15000(RINCOME == 10) + 20000(RIN-
COME == 11) +
25000(RINCOME == 12) ) %>% mutate(lnincome = log(income + 1))

data <- data %>% mutate(house_income = 500(INCOME == 1) + 1000(INCOME == 2) + 3000(INCOME
== 3) + 4000(INCOME == 4) + 5000(INCOME == 5) + 6000(INCOME == 6) + 7000(INCOME == 7)
+ 8000(INCOME == 8) + 10000(INCOME == 9) + 15000(INCOME == 10) + 20000(INCOME == 11) +
25000(INCOME == 12) ) %>% mutate(lnhouse_income = log(house_income + 1))

data <- data %>% mutate(partner_income = house_income - income) %>% mutate(lnpartner_income =
log(partner_income + 1))

data <- data %>% mutate(belief = 1(HELL == 1)(HEAVEN == 1))

mdata <- data %>% filter(SEX == 1) fdata <- data %>% filter(SEX == 2)

```

Age distribution

```

ggplot(data, aes(x = AGE)) + geom_histogram(binwidth = 5, fill = "skyblue", color = "black") + labs(title
= "Age Distribution", x = "Age", y = "Frequency")

```

Education distribution

```

ggplot(data, aes(x = EDUC)) + geom_density(fill = "skyblue", color = "black") + labs(title = "Education
Distribution in Years", x = "Years of Education", y = "Density")

```

Gender distribution

```

ggplot(data, aes(x = factor(SEX), fill = factor(SEX))) + geom_bar() + scale_fill_manual(values = c("1"
= "lightblue", "2" = "pink")) + labs(title = "Gender Distribution", x = "Gender", y = "Count") +
scale_x_discrete(labels = c("1" = "Male", "2" = "Female"))

```

Respondent Income distribution

```

ggplot(data, aes(x = RINCOME)) + geom_histogram(binwidth = 1, fill = "skyblue", color = "black") +
labs(title = "Respondent Income Distribution", x = "RIncome", y = "Frequency")

ggplot(data, aes(x = income)) + geom_histogram(binwidth = 1000, fill = "skyblue", color = "black") +
labs(title = "Respondent Income Distribution", x = "Income", y = "Frequency")

```

Family Income distribution

```

ggplot(data, aes(x = INCOME)) + geom_histogram(binwidth = 1, fill = "skyblue", color = "black") +
labs(title = "Household Income Distribution", x = "Income", y = "Frequency")

```

```
ggplot(data, aes(x = house_income)) + geom_histogram(binwidth = 1000, fill = "skyblue", color = "black")
+ labs(title = "Household Income Distribution", x = "House Income", y = "Frequency")
```

Partner Income distribution

```
ggplot(data, aes(x = partner_income)) + geom_histogram(binwidth = 1000, fill = "skyblue", color =
"black") + labs(title = "Partner Income Distribution", x = "Partner Income", y = "Frequency")
```

Health distribution

```
ggplot(data, aes(x = HEALTH)) + geom_histogram(binwidth = 1, fill = "skyblue", color = "black") +
labs(title = "Health Distribution", x = "Health", y = "Frequency")
```

Happyness distribution

```
ggplot(data, aes(x = HAPPY)) + geom_histogram(binwidth = 1, fill = "skyblue", color = "black") +
labs(title = "Happyness Distribution", x = "Happyness", y = "Frequency")
```

Econometric models (Regression models)

Health with personal income

```
model_health_income <- lm(HEALTH ~ income, data)

summary(model_health_income) modelsummary(model_health_income, gof_omit = "Log.Lik.|AIC|BIC|F|RMSE",
stars = TRUE, notes = "Standard errors in parentheses.", title = "health ~ income")

ggplot(data, aes(x = income, y = HEALTH)) + geom_point() + geom_smooth(method = "lm", se = FALSE,
col = "blue") + labs(title = "Regression Plot of HEALTH ~ income", x = "income", y = "HEALTH")
```

Happy with Health and income

```
model_happy_health_income <- lm(HAPPY ~ HEALTH + income + AGE + EDUC, data)

summary(model_happy_health_income) modelsummary(model_happy_health_income, gof_omit =
"Log.Lik.|AIC|BIC|F|RMSE", stars = TRUE, notes = "Standard errors in parentheses.", title = "happy ~
health + income + AGE + EDUC")

ggplot(data, aes(x = HEALTH + income + AGE + EDUC , y = HAPPY)) + geom_point() +
geom_smooth(method = "lm", se = FALSE, col = "blue") + labs(title = "Regression Plot of HAPPY ~
HEALTH + income", x = "HEALTH + income + AGE + EDUC", y = "HAPPY")

ggplot(data, aes(x = HEALTH, y = HAPPY)) + geom_point() + geom_smooth(method = "lm", se =
FALSE, col = "blue") + labs(title = "Regression Plot of HAPPY ~ RINCOME", x = "HEALTH", y =
"HAPPY")
```

Personal education with father/mother educatione

```
model_edu <- lm(EDUC ~ MAEDUC + PAEDUC + SPEDUC, data)

summary(model_edu) modelsummary(model_edu, gof_omit = "Log.Lik.|AIC|BIC|F|RMSE", stars = TRUE,
notes = "Standard errors in parentheses.", title = "education ~ fother/mother/spouse education")
```

```
ggplot(data, aes(x = MAEDUC + PAEDUC + SPEDUC, y = EDUC)) + geom_point() +
geom_smooth(method = "lm", se = FALSE, col = "blue") + labs(title = "Regression Plot of
RINCOME ~ INCOME", x = "MAEDUC + PAEDUC + SPEDUC", y = "EDUC")
```

Personal income and partner income

```
model_income_partner <- lm(partner_income ~ income, data)
```

```
summary(model_income_partner) modelsummary(model_income_partner, gof_omit = "Log.Lik.|AIC|BIC|F|RMSE",
stars = TRUE, notes = "Standard errors in parentheses.", title = "partner_income ~ income")
```

```
ggplot(data, aes(x = income, y = partner_income)) + geom_point() + geom_smooth(method = "lm", se =
FALSE, col = "blue") + labs(title = "Regression Plot of partner_income ~ income", x = "income", y =
"partner_income")
```

```
model_income_partner_log <- lm(lnpartner_income ~ lnincome, data)
```

```
summary(model_income_partner_log) modelsummary(model_income_partner_log, gof_omit =
"Log.Lik.|AIC|BIC|F|RMSE", stars = TRUE, notes = "Standard errors in parentheses.", title = "Log
partner_income ~ Log income")
```

```
ggplot(data, aes(x = lnincome, y = lnpartner_income)) + geom_point() + geom_smooth(method = "lm",
se = FALSE, col = "blue") + labs(title = "Regression Plot of lnpartner_income ~ income", x = "log income",
y = "log partner_income")
```

income and AGE, EDUC, BABIES male and female data

```
models <- list() models[["male_level"]] <- lm(income ~ AGE + EDUC + BABIES, mdata) mod-
els[["male_log"]] <- lm(lnincome ~ AGE + EDUC + BABIES, mdata) models[["female_level"]] <- lm(income
~ AGE + EDUC + BABIES, fdata) models[["female_log"]] <- lm(lnincome ~ AGE + EDUC + BABIES,
fdata) models[["both"]] <- lm(lnincome ~ AGE + EDUC + BABIES, data) models[["both_log"]] <-
lm(lnincome ~ AGE + EDUC + BABIES, data)
```

```
modelsummary(models, gof_omit = "Log.Lik.|AIC|BIC|F|RMSE", stars = TRUE, notes = "Standard errors
in parentheses.", title = "Income table")
```

```
ggplot(data, aes(x = AGE, y = income)) + geom_point() + geom_smooth(method = "lm", se = FALSE,
col = "blue") + labs(title = "Regression Plot of income ~ AGE", x = "AGE", y = "income")
```

```
ggplot(mdata, aes(x = AGE, y = income)) + geom_point() + geom_smooth(method = "lm", se = FALSE,
col = "blue") + labs(title = "Male Regression Plot of income ~ AGE", x = "AGE", y = "income")
```

```
ggplot(fdata, aes(x = AGE, y = income)) + geom_point() + geom_smooth(method = "lm", se = FALSE,
col = "blue") + labs(title = "Female Regression Plot of income ~ AGE", x = "AGE", y = "income")
```

```
ggplot(data, aes(x = EDUC, y = income)) + geom_point() + geom_smooth(method = "lm", se = FALSE,
col = "blue") + labs(title = "Regression Plot of income ~ EDUC", x = "EDUC", y = "income")
```

```
ggplot(mdata, aes(x = EDUC, y = income)) + geom_point() + geom_smooth(method = "lm", se = FALSE,
col = "blue") + labs(title = "Male Regression Plot of income ~ EDUC", x = "EDUC", y = "income")
```

```
ggplot(fdata, aes(x = EDUC, y = income)) + geom_point() + geom_smooth(method = "lm", se = FALSE,
col = "blue") + labs(title = "Female Regression Plot of income ~ EDUC", x = "EDUC", y = "income")
```

```
ggplot(data, aes(x = BABIES, y = income)) + geom_point() + geom_smooth(method = "lm", se = FALSE,
col = "blue") + labs(title = "Regression Plot of income ~ BABIES", x = "BABIES", y = "income")
```

```
ggplot(mdata, aes(x = BABIES, y = income)) + geom_point() + geom_smooth(method = "lm", se =
FALSE, col = "blue") + labs(title = "Male Regression Plot of income ~ BABIES", x = "BAABIES", y =
"income")
```

```
ggplot(fdata, aes(x = BABIES, y = income)) + geom_point() + geom_smooth(method = "lm", se = FALSE,  
col = "blue") + labs(title = "Female Regression Plot of income ~ BABIES", x = "BABIES", y = "income")
```