Coding Sample

2025-10-11

Uploading the Dataset into R Markdown

```
cmpst_scores <- read_excel("/Users/hamid/Documents/UC Berkeley/URAP/Garden Questionnaire Dataset/GardenQues/Main.xlsx")
```

```
## New names:
## • `` -> `...413`
## • `` -> `...414`
## • `` -> `...415`
## • `` -> `...416`
```

Notes on Cleaning:

This code uses a fairly clean dataset with few missing variables. Therefore, the cleaning part is short and does not includes the common techniques we use with big data

Varilabes Names were uploaded and used in the following codes. However, code was not included since it takes a lot of space.

The code chunk below generates a score that will be used in the composite score. It first makes a variable called LeanMonths_score which calculates the amount of lean month (no harvest) per year and then inverts it by subtracting it from 12 so that higher scores represent more harvests.

```
#Binary numbers for each column is summed and total 12. If a household have 12 months with no harvest, they would get zero for this score because the sum is subtracted from 12.
LeanMonths_score = 12 - rowSums(cmpst_scores[,c("16. January", "16. February", "16. March", "16. April", "16. May", "16. June", "16. July", "16. August", "16. September", "16. October", "16. November", "16. December"  )], na.rm = TRUE)

#this line of code manipulates and combines three quantitative variables including the score generate prior to this, a food security score, and a binary score. Different weight are given to each based on supervisors instructions to each and they total to 10.
cmpst_scores = cmpst_scores %>% mutate(food_security = (LeanMonths_score/3)+(`HFIAS Inverse Score` * 4 / 21) +
      ifelse(`30. Do you grow vegetables?` == "Yes", 2, 0))
```

The code chunk below generate the second score (resilience score) for the composite score. It is made up of 5 sub scores that sum to 20. Resilience is the most important contributer to the final composite score.

```
#The line below sums households access to different sources of water
water_sources = rowSums(cmpst_scores[,c("11. No water access", "11. River/Creek/Stream",
"11. Rainwater Harvest", "11. Spring Water", "11. Well Water", "11. Borehole/Groundwate
r", "11. Tap/Piped")],na.rm = TRUE)

#The line below sums the numbers of essential assets households have
assets_house = rowSums(cmpst_scores[,c("14. Improved Stove", "14. Refrigerator", "14. Mo
bile phone", "14. Smart mobile phone", "14. Bicycle", "14. Motorcycle", "14. Radio", "1
4. Television")],na.rm = TRUE)

#The line below sums the total numbers of illnesses a household have. It will be subtrac
ted from 8 later.
illnesses = rowSums(cmpst_scores[,c("Malaria", "Respiratory diseases/cough", "Diarrheal
diseases", "Tuberculosis", "HIV-AIDS", "Cholera", "Chronic (diabetes, heart disease, can
cer, hypertension/blood pressure)", "Ulcers")],na.rm = TRUE)

#The line below sums the diverse sourves of incomes for each household.
income_sources=rowSums(cmpst_scores[,c("65. Day labor wages", "65. Sales of vegetables f
rom my garden", "65. Farm sales - food", "65. Farm sales - commercial", "65. Farm sales
- animals", "65. Small business")],na.rm = TRUE)

#This line sums the number of diverse farming practices a household uses.
vegetable_practice = rowSums(cmpst_scores[,c("Nursery bed preparation", "Transplanting t
o raised beds", "Use of compost", "Use of natural/botanical pesticide", "Mulching", "Com
panion/intercropping", "Seed removal/saving", "Rainwater harvesting")],na.rm = TRUE)

#Finally all the codes based on the given instructions are added with their weights in t
he code below.
cmpst_scores$resilience_score = (water_sources*4/7) + (assets_house*3/8) + ((8-illnesse
s)/2) + (income_sources) + (vegetable_practice/2)
```

The code chunk below makes the final score in the composite score and is called Women Empowerment Score. It is more qualitative in nature. The codes below converts their qualitative response to binary and then adds them up at the end

```
#The code below turns 8 variables into binaries which add up to 14. In the end, they are combined and the weights are adjusted to that it adds to 10.
cmpst_scores <- cmpst_scores %>%
  mutate(
    house_ownership_score = case_when(
      `3. Do you own this or any other house either alone or jointly with someone else?` == "Alone" ~ 2,
      `3. Do you own this or any other house either alone or jointly with someone else?` == "Jointly" ~ 1,
      TRUE ~ 0
    ),
    farmland_ownership_score = case_when(
      `4. Do you own farmland either alone or jointly with someone else?` == "Alone" ~ 2,
      `4. Do you own farmland either alone or jointly with someone else?` == "Jointly" ~ 1,
      TRUE ~ 0
    ),
    title_deed_score = ifelse(`6. Do you have a title deed for any land you own?` == "Yes", 1, 0),
    name_on_deed_score = ifelse(`7. Is your name on the title deed?` == "Yes", 1, 0),
    decision_on_income_score = case_when(
      `62. Who decides how the garden money is used?` == "I decide on my own how to spend the garden income" ~ 2,
      `62. Who decides how the garden money is used?` == "I discuss how to spend it with my husband/partner" ~ 1,
      TRUE ~ 0
    ),
    confidence_score = case_when(
      `81. “I believe in myself and am confident to take on new things”` == "Yes, definitely" ~ 2,
      `81. “I believe in myself and am confident to take on new things”` == "Yes, somewhat" ~ 1,
      TRUE ~ 0
    ),
    goals_and_dreams_score = case_when(
      `82. “I have a full idea of my future goals and dreams”` == "Yes, definitely" ~ 2,
      `82. “I have a full idea of my future goals and dreams”` == "Yes, somewhat" ~ 1,
      TRUE ~ 0
    ),
    proud_of_self_score = case_when(
      `83. “On the whole, I feel proud of myself”` == "Yes, definitely" ~ 2,
      `83. “On the whole, I feel proud of myself”` == "Yes, somewhat" ~ 1,
      TRUE ~ 0
    ),
    womens_empowerment_score = (coalesce(house_ownership_score, 0) +
      coalesce(farmland_ownership_score, 0) +
      coalesce(title_deed_score, 0) +
      coalesce(name_on_deed_score, 0) +
      coalesce(decision_on_income_score, 0) +
      coalesce(confidence_score, 0) +
      coalesce(goals_and_dreams_score, 0) +
```

```
      coalesce(proud_of_self_score, 0))*5/7
  )
```

The code chunk below combine all the subscores into the composite score which will later be used to compare the Target and Control households

```
#All the subscores are combined and a new variable is made.
cmpst_scores$total_score = cmpst_scores$food_security + cmpst_scores$resilience_score +
cmpst_scores$womens_empowerment_score
```

The code chunk below generates a Line Histogram for all the three subscores and the final composite score in one chart.

```
#A new variable is made called group to distinguish between Target and Control household
s
cmpst_scores$Group=if_else(cmpst_scores$Household_code >= 200, "Target", "Control")

#Line Histogram for the food security
FS_GGp=ggplot(cmpst_scores, aes(x=food_security, color=Group))+
      geom_density(size = 1, trim=TRUE)+
      xlim(0, 10) +
  labs(
    title = "Food Security Score by Group",
    x = "Food Security Score",
    y = "Density"
  ) + theme_bw()
```

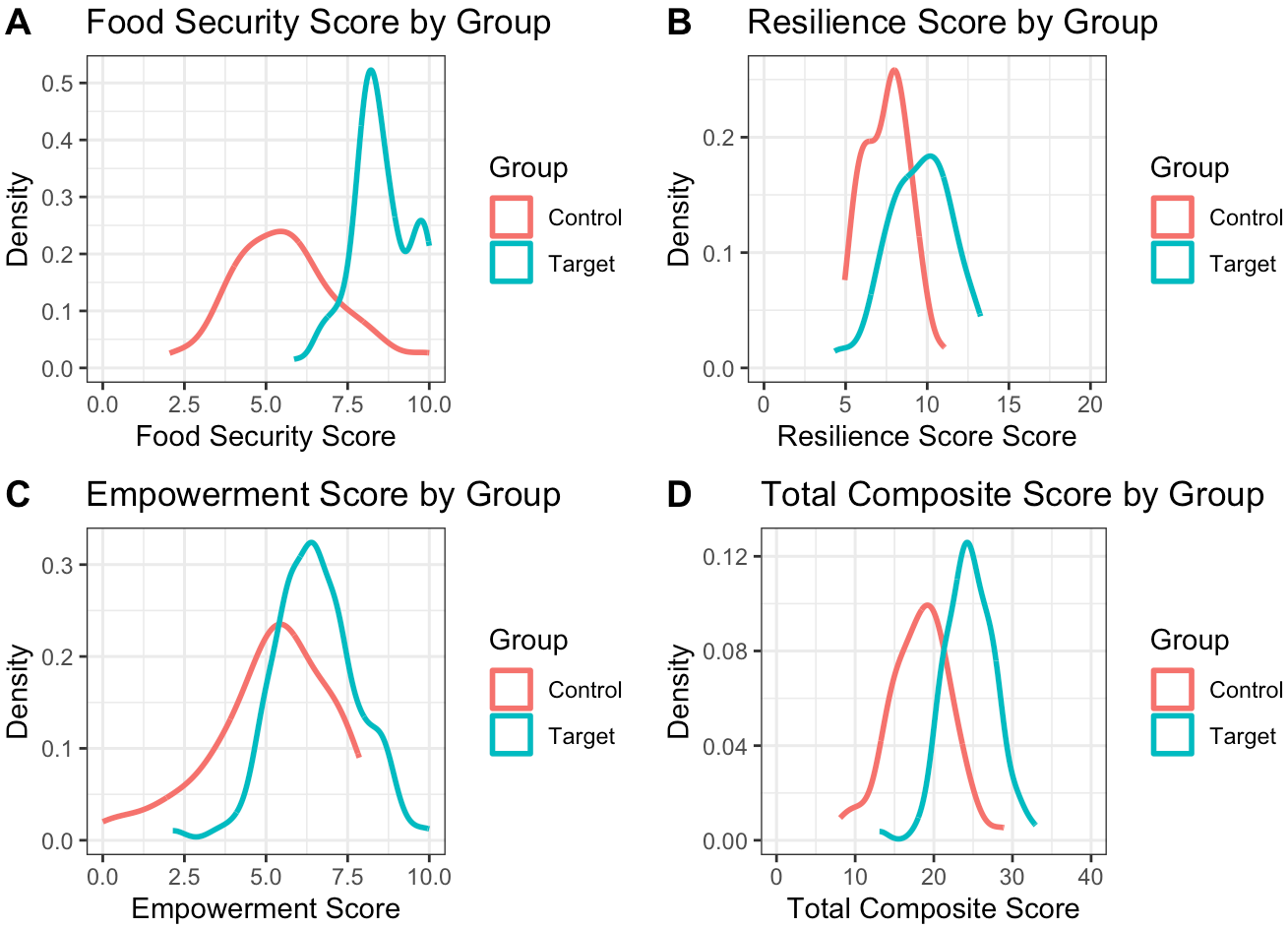
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
#Line Histogram for the resilience score
RS_GGp=ggplot(cmpst_scores, aes(x=resilience_score, color=Group))+
      geom_density(size = 1, trim=TRUE)+
      xlim(0, 20) +
  labs(
    title = "Resilience Score by Group",
    x = "Resilience Score Score",
    y = "Density"
  ) + theme_bw()

#Line Histogram for the women empowerment score
WES_GGp=ggplot(cmpst_scores, aes(x=womens_empowerment_score, color=Group))+
      geom_density(size = 1, trim=TRUE)+
      xlim(0, 10)+
  labs(
    title = "Empowerment Score by Group",
    x = "Empowerment Score",
    y = "Density"
  ) + theme_bw()

#Line Histogram for the total score
TS_GGp=ggplot(cmpst_scores, aes(x=total_score, color=Group))+
      geom_density(size = 1, trim=TRUE)+
      xlim(0, 40)+
  labs(
    title = "Total Composite Score by Group",
    x = "Total Composite Score",
    y = "Density"
  ) + theme_bw()

#All the plots combined in a grid
plot_grid(FS_GGp, RS_GGp, WES_GGp, TS_GGp,
          labels = c("A", "B", "C", "D"),
          ncol = 2, nrow = 2)
```



This code chunks puts the number for the total score in a quartile table grouping by Control and Target

```
#This line groups the households into target and control and then summarises the total s
core by generating quartiles, mean, and SD.
compst_scores_stats <- as.data.frame(cmpst_scores %>%
  group_by(Group) %>%
  summarise(
    Min = min(total_score, na.rm = TRUE),
    First_Quartile = quantile(total_score, 0.25, na.rm = TRUE),
    Median = median(total_score, na.rm = TRUE),
    Third_Quartile = quantile(total_score, 0.75, na.rm = TRUE),
    Max = max(total_score, na.rm = TRUE),
    Mean = mean(total_score, na.rm = TRUE),
    SD = sd(total_score, na.rm = TRUE)
  ))

#This code generates a table from the numbers calculated from the prior code.
compst_scores_stats %>%
  gt() %>%
  tab_header(
    title = "Comparison of Total Score Distribution between Target and Control"
  ) %>%
  fmt_number(
    columns = where(is.numeric),
    decimals = 2
  )%>%
  tab_options(
    column_labels.font.weight = "bold"
  )
```

Comparison of Total Score Distribution between Target and Control

Group	Min	First_Quartile	Median	Third_Quartile	Max	Mean	SD
Control	8.09	15.32	18.49	20.66	28.93	18.27	3.97
Target	13.08	22.64	24.51	26.65	32.96	24.58	3.16

This code chunk compares the means for each score between Control and Target and calculate the t state and p value of them.

```
#this puts all the required variables in a vectors.
final_scores <- c("food_security", "resilience_score", "womens_empowerment_score", "tota
l_score")

#The code below runs the t.test function for all the scores above grouping by Target and
Control. Transmute select only the intended results for the table.
t_test_results <- lapply(final_scores, function(var) {
  formula <- as.formula(paste(var, "~Group"))
  test <- t.test(formula, data = cmpst_scores)

  tidy(test)%>%
   transmute(
      Variable = var,
      Target_Mean = estimate2,
      Control_Mean = estimate1,
      Difference = Target_Mean - Control_Mean,
      T_Stat = statistic,
      P_Value = p.value
    )
    })

#Turning the table into dataframe to avoid unnecessary texts that R automaticaly generat
es.
final_table=as.data.frame(bind_rows(t_test_results))

#Final edits to the table are made below
final_table %>%
  gt() %>%
  tab_header(
    title = "Comparison of Means between Target and Control Groups",
    subtitle = "Variables: Food Security, Resilience, Women's Empowerment, Total Score"
  ) %>%
  fmt_number(
    columns = where(is.numeric),
    decimals = 2
  )%>%
  tab_options(
    column_labels.font.weight = "bold"
  )
```

Comparison of Means between Target and Control Groups

Variables: Food Security, Resilience, Women's Empowerment, Total Score

Variable	Target_Mean	Control_Mean	Difference	T_Stat	P_Value
food_security	8.51	5.65	2.86	−11.11	0.00
resilience_score	9.56	7.51	2.05	−7.39	0.00
womens_empowerment_score	6.51	5.11	1.40	−4.86	0.00

Comparison of Means between Target and Control Groups

Variables: Food Security, Resilience, Women's Empowerment, Total Score

Variable	Target_Mean	Control_Mean	Difference	T_Stat	P_Value
total_score	24.58	18.27	6.31	−9.80	0.00