Project 1 Peer Review

REGAN (Will, Nathaniel, Dinu, and Muhammad)

Review Group Name: Team Hokies

Answer the following about the other group.

Response Variable

- What is the response variable? Median Household Income
- Based on the histogram does it look appropriate? Describe the shape of the distribution. It appears to be unimodal. It appears to be symmetric and looks appropriate.

Research Questions

- What are the three variables of interest?

1) Population
2) Percentage of People who have a High School Degree
3) Average Size Home

- What suggestions do you have to improve the specificity of these questions?

1) Perhaps looking specifically at different locations within a state.
2) Address possible sources of biases or confounding.
3) Analyze what variables of interest make the MOST SENSE with this response variables.

Data and Variables

- list the primary sources used

The primary source was PolicyMap. Also, Statista was used for average household size from the US Government Census.

- Is the data in one single sheet?

All of the data is in a single data sheet.

- What is an observation (row) in the data?

The observations (rows) are all states.

- Is “time” a factor (year, month, season, etc)?

Time is not a variables in their data.

- Are there at least 3 qualitative and 3 quantitative. And a total of 8 variables?

Quantitative: 1) Size 2) Price of College 3) Population

Qualitative: 1) Region 2) Urban vs. Rural 3) Political

- For qualitative variables, are there less than 5 levels for each?

They do not have more than five levels. Political is 3 levels, region is 4, and Urban vs. Rural is 2.

Overall comments

- What suggestions do you have moving forward for this group?

1) They need to carefullly consider the variable levels and how those will work.
2) They should consider other variables that may deem more important both socially and economically.

- What improvements might you use in your own project?

1) Discuss with their team what data variables them deem most important and viable.
2) Talk as a team and build comradery for effective workign environemnts.

Your Group Name: REGAN

Answer the following about the your own group.

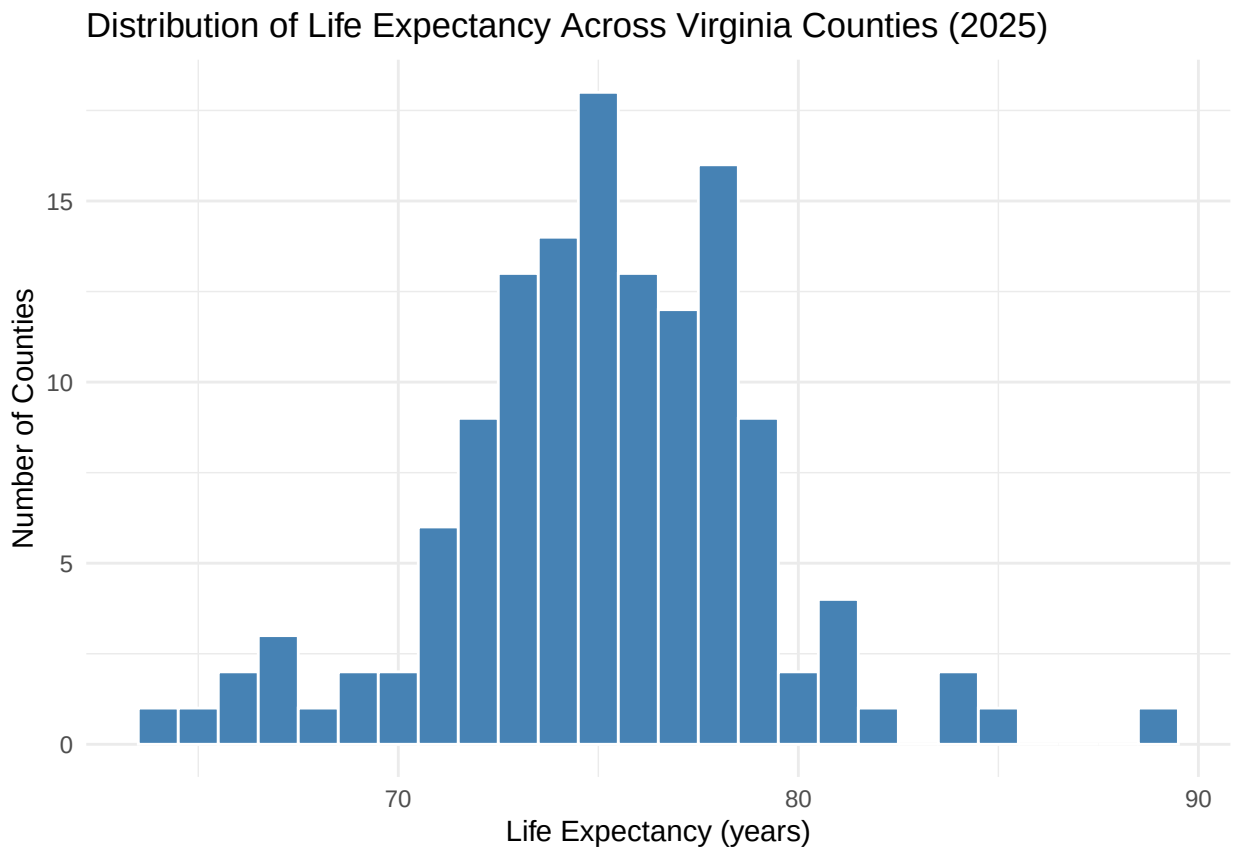
Response Variable

```
#Include a histogram of your response
addl_df <- addl_df %>%
  filter(!is.na(county) & county != "NA")
summary(addl_df$life_expectancy)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  64.32   73.14   75.24   75.15   77.62   88.91       1
```

```
ggplot(addl_df, aes(x = life_expectancy)) +
  geom_histogram(binwidth = 1, fill = "steelblue", colour = "white") +
  labs(title = "Distribution of Life Expectancy Across Virginia Counties (2025)",
       x     = "Life Expectancy (years)",
       y     = "Number of Counties") +
  theme_minimal()
```

```
Warning: Removed 1 row containing non-finite outside the scale range
(`stat_bin()`).
```



- What is the response variable?

Life Expectancy (Years)

- Based on the histogram does it look appropriate? Describe the shape of the distribution. What is the range of values?

It looks appropriate based on the shape. It is unimodal and mostly symmetric. The range of data is from 64.32 to 88.91 (24.59). Also, our mean and median are close to the national average life expectancy.

- What comments did your review group provide?

1) They should we should compile all of the data we need into excel sheet which we largely agree with, as our data is currently in 6 different sheets in one excel book.
2) They should we should try to not account for every possible variable in our research questions as this can be tedious.
3) They said we should define our variables numerically and what they mean (i.e. physician ratio)

```
#Include any other plots you may have

#No other plots at this time
```

Research Questions

- List your three research questions (include any corrections or comments from your reviewers.

Socio-economic factors: 1) To what extent do median household income and income inequality jointly explain county-level life expectancy?

Health-care access: 2) Is a lower primary-care physician ratio associated with higher life expectancy when controlling for the overall health-infrastructure tier and adult obesity prevalence?

Behavioral and Environmental Conditions: 3) Do counties with lower adult obesity and greater access to exercise opportunities exhibit longer life expectancy, independent of particulate air pollution levels?

Data and Variables

```
#include the head() of your data table
head(addl_df) #sheet 4 of our excel book
```

```
# A tibble: 6 x 395
  fips  state  county life_expectancy x95_percent_ci_low_5 x95_percent_ci_high_6
  <chr> <chr>  <chr>            <dbl>                <dbl>                 <dbl>
1 51001 Virgi~ Accom~            73.8                 72.8                  74.9
2 51003 Virgi~ Albem~            81.5                 80.9                  82.0
3 51005 Virgi~ Alleg~            73.1                 71.5                  74.7
4 51007 Virgi~ Amelia            73.6                 72.0                  75.2
5 51009 Virgi~ Amher~            75.5                 74.5                  76.5
6 51011 Virgi~ Appom~            77.2                 75.8                  78.5
# i 389 more variables: life_expectancy_hispanic_all_races <dbl>,
```

```
#   life_expectancy_hispanic_all_races_95_percent_ci_low <dbl>,
#   life_expectancy_hispanic_all_races_95_percent_ci_high <dbl>,
#   life_expectancy_non_hispanic_aian <lgl>,
#   life_expectancy_non_hispanic_aian_95_percent_ci_low <lgl>,
#   life_expectancy_non_hispanic_aian_95_percent_ci_high <lgl>,
#   life_expectancy_non_hispanic_asian <dbl>, ...
```

- List the primary sources used

Countyhealthrankings.org: https://www.countyhealthrankings.org/health-data/virginia?year=2025&measure=Population+Health+and+Well-being&mapView=state

- Is the data in one single sheet?

The data is not one sheet. It is six different sheets.

- What is an observation (row) in the data?

An observation (row) in the data is a single county.

- Is time a factor?

While our repsonse variable is measured in years, we are only looking at a single value in the year 2025, so our data is not adjusted over time.

- Do you meet the criteria for the explanatory variable?

Yes, we meet the criteria stated in the rubric for the explanatory variable.

- Were there any suggestions about the variables you might incorporate?

While they did not suggest any new variables, they said we should try to define what they are and how they are calculated.

Overall comments

- From this session, what improvements might you use in your project?

1) We need to definitely get a better grasp of our variables and what they do.
2) We need to look at all of the

- What is your group's plan to successfully complete this project moving forward?

1) Work diligently every day.
2) Set clear goals and research expectations.
3) Communicate effectively in Imessage and be aware of other teammates.