HW4

Miracle Ephraim

Question 1

Remove all SNPs, 800-series plans, and prescription drug only plans (i.e., plans that do not offer Part C benefits). Provide a box and whisker plot showing the distribution of plan counts by county over time. Do you think that the number of plans is sufficient, too few, or too many?

```
full_df %>%
  group_by(fips, year) %>%
  select(fips, year) %>%
  summarize(plan_count = n()) %>%
  ggplot(aes(x = as.factor(year), y = plan_count)) +
  geom_boxplot() +
  labs(
       x = "Year",
       y = "Count of Plans"
  ) +
  scale_y_continuous(breaks = seq(0, 0), limits = c(0, 40)) +
  theme_bw()
```

Question 2

Provide bar graphs showing the distribution of star ratings in 2010, 2012, and 2015. How has this distribution changed over time?

```
for (y in c(2010, 2012, 2025)) {
  full_df %>%
    filter(year == y) %>%
    filter(Star_Rating != "NA") %>%
  ggplot(aes(x=as.factor(Star_Rating))) +
  geom_bar() +
  labs(
    x="Star Rating",
    y="Count of Plans",
    title=paste("Frequency Distribution of Star Ratings in ",y)
  ) + theme_bw()
}
```

Question 3

Plot the average benchmark payment over time from 2010 through 2015. How much has the average benchmark payment risen over the years?

```
ggplot(q3, aes(x = year, y = avg_rate)) +
    geom_line() +
    geom_point() +
    labs(
        title = "Average MA Rate Over Time",
        x = "Year",
        y = "Average MA Rate"
    ) +
    theme_minimal()
```

Question 4

Plot the average share of Medicare Advantage (relative to all Medicare eligibles) over time from 2010 through 2015. Has Medicare Advantage increased or decreased in popularity? How does this share correlate with benchmark payments?

```
ggplot(q4_b, aes(x = year, y = avg)) +
    geom_line() +
    geom_point() +
    labs(
        title = "Average Share of Enrollment to Eligible Population Over Time",
        x = "Year",
        y = "Average Share"
    ) +
    theme_minimal()
```

For the rest of the assignment, we'll use a regression discontinuity design to estimate the average treatment effect from receiving a marginally higher rating. We'll focus only on 2010.

Question 5

Calculate the running variable underlying the star rating. Provide a table showing the number of plans that are rounded up into a 3-star, 3.5-star, 4-star, 4.5-star, and 5-star rating.

```
print(star_rating_summary)
```

Question 6

Using the RD estimator with a bandwidth of 0.125, provide an estimate of the effect of receiving a 3-star versus a 2.5 star rating on enrollments. Repeat the exercise to estimate the effects at 3.5 stars, and summarize your results in a table.

```
rd.est <- rdrobust(
  y = df_10c$avg_enrollment,
  x = df_10c$Star_Rating,
  c = 3,
  h = 0.125
)

rd.est <- rdrobust(
  y = df_10c$avg_enrollment,
  x = df_10c$Star_Rating,
  c = 3,
  h = 0.125
)
```

Question 7

Repeat your results for bandwidhts of 0.1, 0.12, 0.13, 0.14, and 0.15 (again for 3 and 3.5 stars). Show all of the results in a graph. How sensitive are your findings to the choice of bandwidth?

```
print(plot_3_vs_2_5)
print(plot_3_5_vs_3)
```

Question 8

Examine (graphically) whether contracts appear to manipulate the running variable. In other words, look at the distribution of the running variable before and after the relevent threshold values. What do you find?

Question 9

Similar to question 4, examine whether plans just above the threshold values have different characteristics than contracts just below the threshold values. Use HMO and Part D status as your plan characteristics.

```
print(characteristics_3_vs_2_5)

print(characteristics_3_5_vs_3)
```

Question 10

Summarize your findings from 5-9. What is the effect of increasing a star rating on enrollments? Briefly explain your results.