

1. Evaluate Graphic from Journal News

a. Comparisons

This graph compares the gas price per gallon in the United States over Memorial Day weekend for years 2005 to 2017. The comparison is shown with a bar chart, where higher bars means higher gas prices. This creator of this graphic also chose to explicitly compare the gas price of 2017 with that of 2016 to point out the most recent change in the observational data. This comparison is shown in the form of subtitle for the graphic, indicating that the creator is most interested in this specific comparison.

b. Causality, Mechanism, Structure, Explanation

In my opinion, the graphic does not show that change in time (years) causes an increase or decrease in gas prices of the holiday weekend. Although the price increase during the middle years shown, the price eventually returns to the original levels, showing no significant change in the price of gas. I do not believe that this graphic does a good job of showing any underlying mechanism or structure of the the change in gas prices relative to Memorial Day weekend. The data might be more meaningful if it showed the prices outside of the weekend to see how the holiday itself affects prices. Despite a small scope of the underlying structure, this graph does show that there was some other factor in the middle years that causes the prices to rise compared to the beginning and ending years.

c. Multivariate Analysis

This graphic clearly includes a multivariate analysis: year vs. gas price of Memorial Day weekend. The year variable on the x-axis immediately shows the audience that we should pay attention to the change in some other variable over time. The multivariate analysis is easily seen as the blue bars showing the price of gas for that year.

d. Integration of Evidence

The graphic has multiple instances of integration of evidence. The first being just the time series data, the focal point of the display. However, the author also chose to include other evidence alongside the main display. The first is the subtitle sentence that points out the exact change in gas price over the last year. Connected to that, the creator also explicitly shows the exact price of gas for 2017 while no other data points like this are shown.

e. Documentation

The creator clearly documents the source of the data at the bottom of the graphic: "US Energy Information Association". This gives the audience a sense of confidence in the validity of the data given that it was obtained by the national government. This gives the evidence shown much more weight in the view of the audience.

f. Content Counts Most of All

The content is indeed the focal point of this graphic. The principle of the data ink ratio is respected and there's not much chart junk to cloud the real message the creator is trying to send. This is reinforced by the fact that the only two colors used are black and blue, with the blue only representing the gas prices which are the main data points. This focusses the attention on what really matters in the graphic instead of other non-critical data.

2. Life expectancy vs. percapita GDP in in 2007

```
# Author: Brian Lambert
# Name: module_3_homework_code.R
# Description: Graphics that compare the distribution of
# percapita GDP in the different continents in 2007.

# setwd("/Users/brianlambert/Desktop/STA404/Module_2")

options(warn = 1)

#===== Data =====
gap_2007 <- gapminder[gapminder$year==2007,]

#===== Graphic A =====
ggplot(gap_2007, aes(x = lifeExp, y = gdpPercap)) +
  geom_jitter(alpha = 0.6, aes(col = continent, size = pop)) +
  labs(title = "life expectancy vs. percapita GDP in 2007",
       x = "life expectancy (years)",
       y = "percapita GDP",
       caption = "Source: Gapminder dataset") +
  theme_minimal()

#===== Graphic B =====
ggplot(gap_2007, aes(x = lifeExp, y = gdpPercap)) +
  geom_jitter(alpha = 0.6, aes(col = continent, size = pop)) +
  geom_smooth(aes(x = lifeExp, y = gdpPercap), method="loess", se=F, alpha = 0.4) +
  labs(title = "life expectancy vs. percapita GDP in 2007",
```

```

x = "life expectancy (years)",
y = "percapita GDP",
caption = "Source: Gapminder dataset") +
theme_minimal()

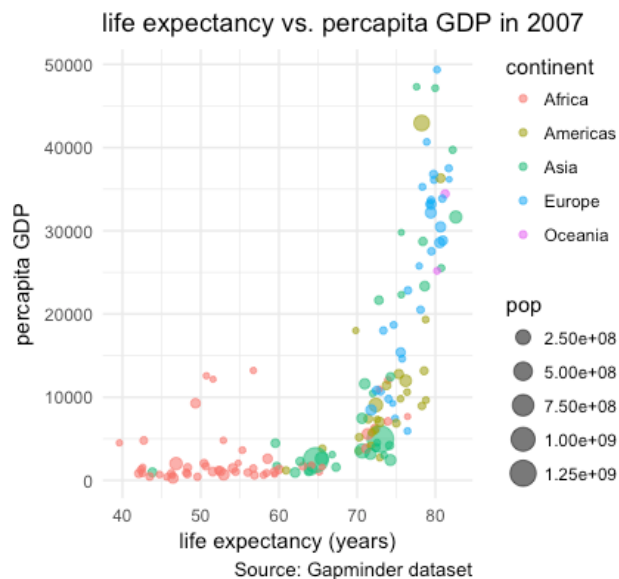
```

```

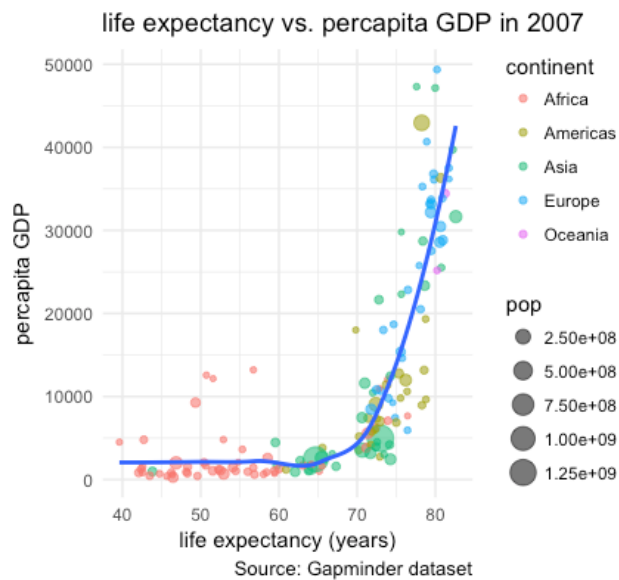
#===== Graphic C =====
ggplot(gap_2007, aes(x = lifeExp, y = gdpPercap)) +
  geom_jitter(alpha = 0.6, aes(col = continent, size = pop)) +
  geom_smooth(aes(col = continent), method="loess", se=F, span = 1) +
  labs(title = "life expectancy vs. percapita GDP in 2007",
       x = "life expectancy (years)",
       y = "percapita GDP",
       caption = "Source: Gapminder dataset") +
  theme_minimal()

```

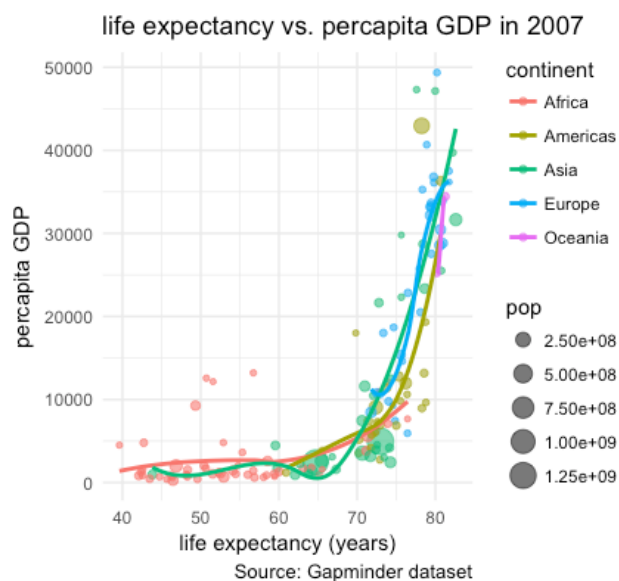
a.



b.



c.



d. Yes, I believe the inclusion of the curve aids audience in seeing that the data follows an exponential trend. It's not difficult to see the trend with the points alone but the addition of the curve simply shines more light on the point. However, this is only my opinion with regards to plot B with the single smooth line. This is a clean overlay that highlights the trend without adding too much complexity to the graph. On the other hand, I believe plot C with the curves for each of the continents masks the data and would only have a negative impact on the audience's understanding.

