Homework 2

Brandon Leslie

Homework 2

Exercise 1

** a). What does the following code do? Does it work? Does it make sense? Why/why not?**

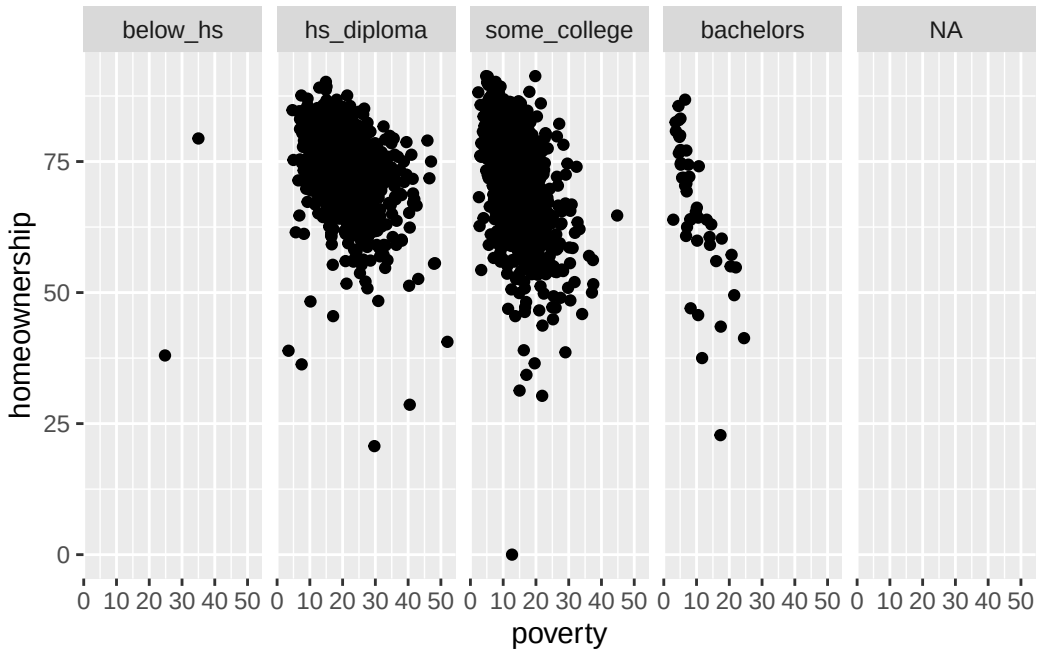
```
ggplot(county) +
  geom_point(aes(x = median_edu, y = median_hh_income)) +
  geom_boxplot(aes(x = smoking_ban, y = pop2017))
```

This code visulizes the relationship between median income & median education. It shows this with the use of boxplots & scatter plots.

** b). Recreate each of the following. Which of these makes it easier to compare poverty levels (poverty) across people from different median education levels (median_edu)? What does this say about when to place a faceting variable across rows or columns? **

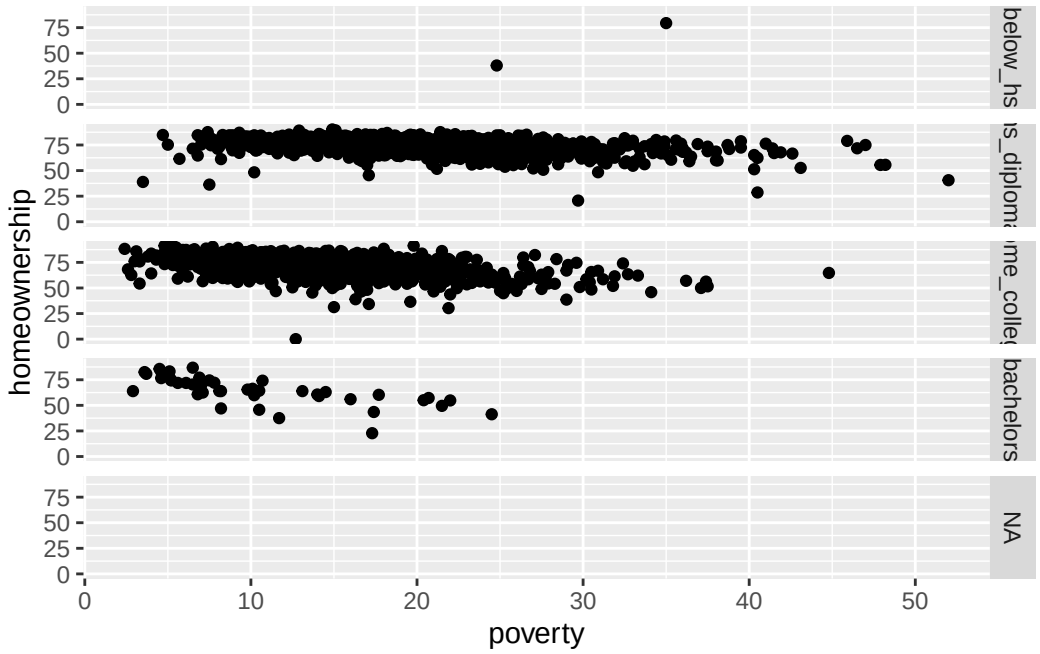
Graph 1

```
county %>%
  ggplot(aes(y = homeownership, x = poverty)) +
  geom_point() +
  facet_grid(.~median_edu)
```



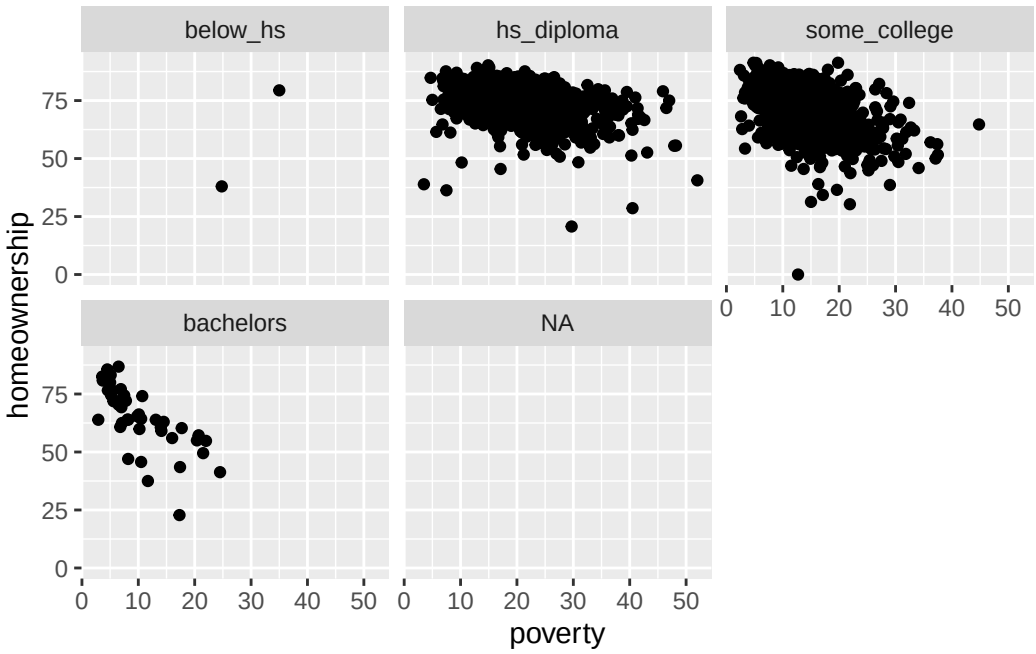
Graph 2

```
county %>%
  ggplot(aes(y = homeownership, x = poverty)) +
  geom_point() +
  facet_grid(median_edu ~ .)
```



Graph 3

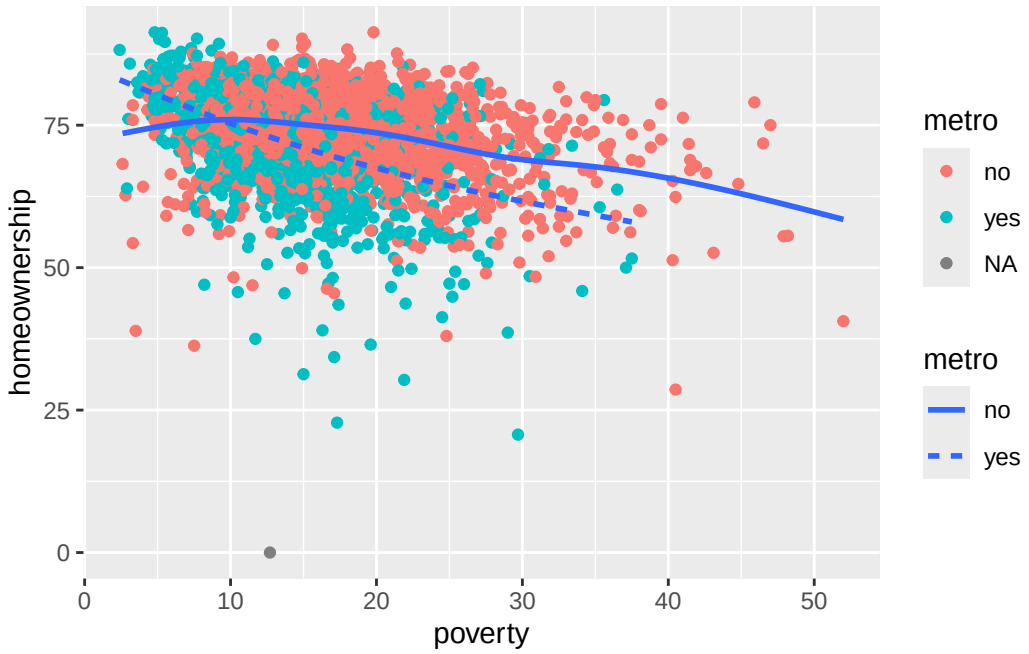
```
county %>%
  ggplot(aes(y = homeownership, x = poverty)) +
  geom_point() +
  facet_wrap(~median_edu)
```



** c).Recreate the following plots.**

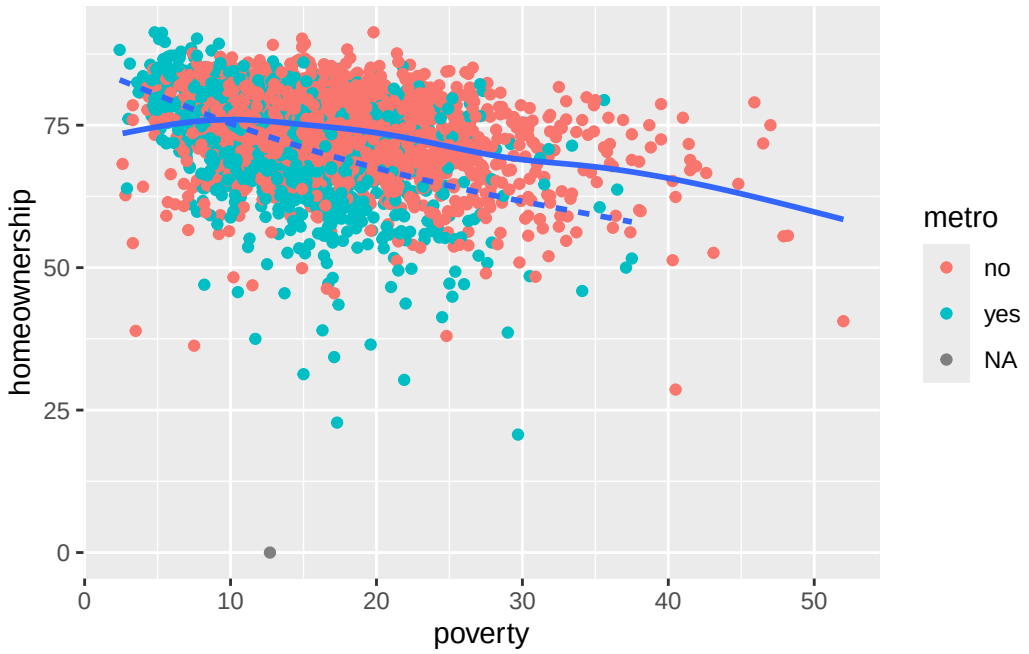
Graph 1

```
county %>%
  ggplot(aes(y = homeownership, x = poverty)) +
  geom_point(aes(color = metro)) +
  geom_smooth(aes(linetype = metro), se = FALSE)
```



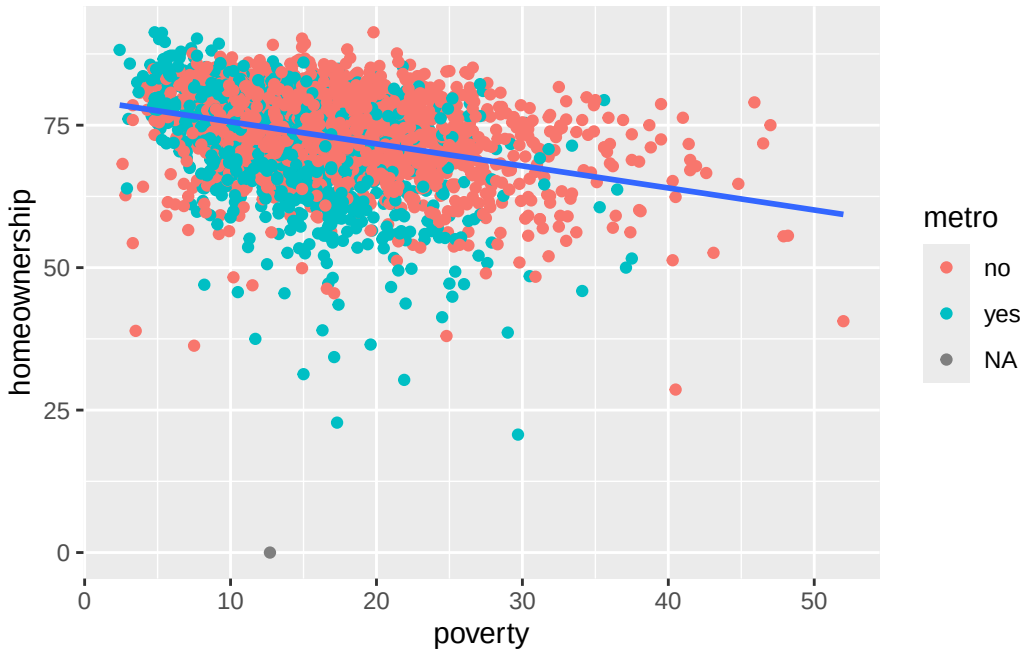
Graph 2

```
county %>%
  ggplot(aes(y = homeownership, x = poverty)) +
  geom_point(aes(color = metro)) +
  geom_smooth(aes(linetype = metro), se = FALSE, show.legend = FALSE)
```



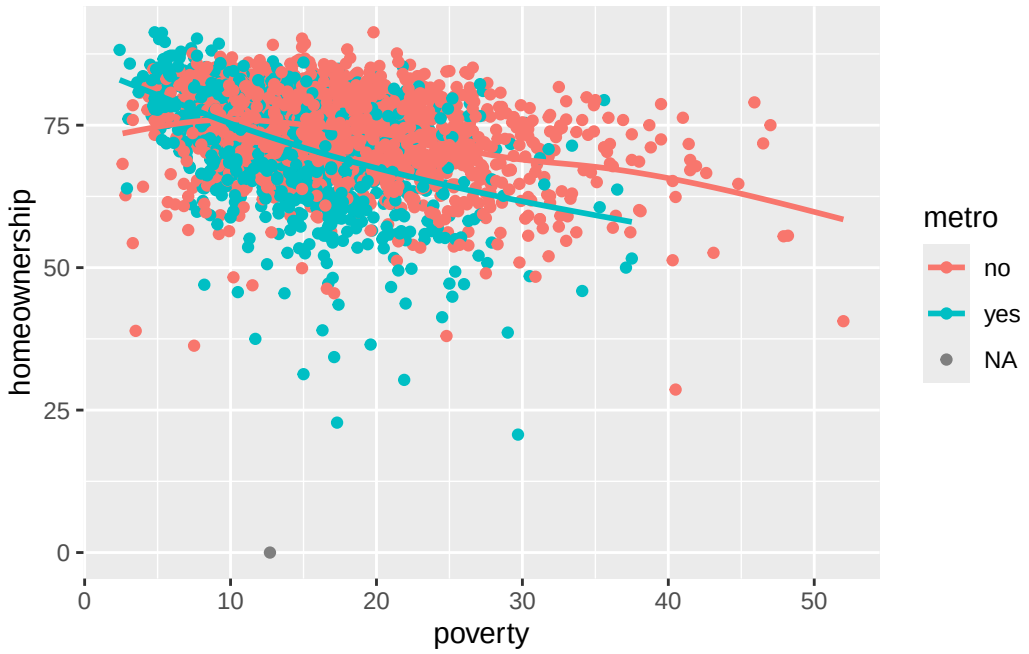
Graph 3

```
county %>%
  ggplot(aes(y = homeownership, x = poverty)) +
  geom_point(aes(color = metro)) +
  geom_smooth(se = FALSE, show.legend = FALSE, method = lm)
```



Graph 4

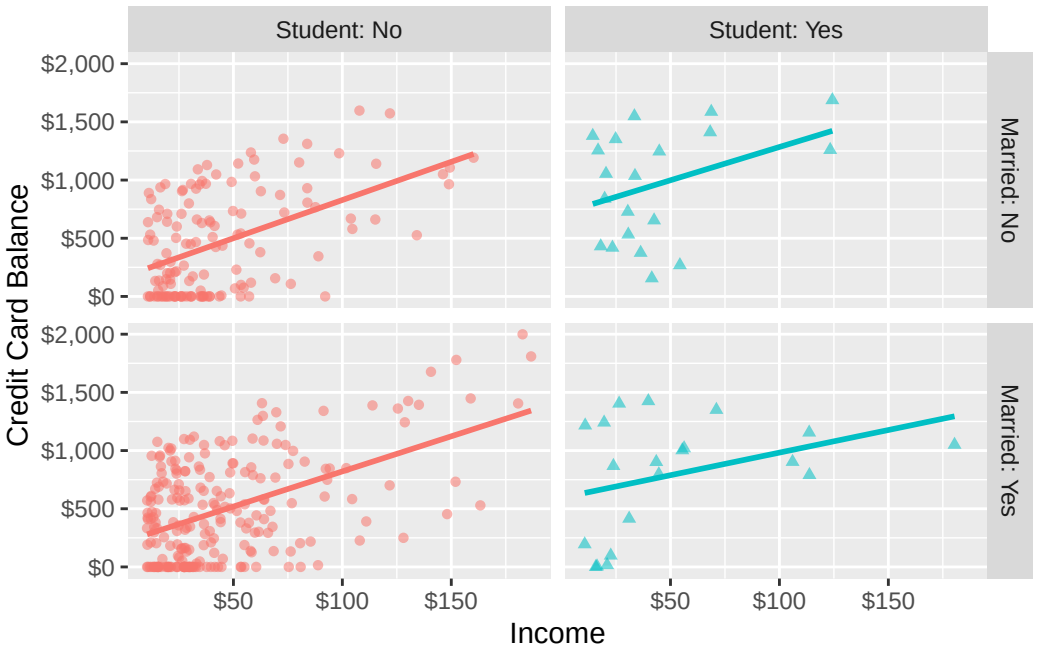
```
county %>%
  ggplot(aes(y = homeownership, x = poverty)) +
  geom_point(aes(color = metro)) +
  geom_smooth(aes(color = metro), se = FALSE, show.legened = FALSE)
```



Exercise 2

a. Recreate the following visualization.

```
Credit %>%
  ggplot(aes( y = Balance, x = Income, color = Student, shape = Student, show.legend = FALSE)
  geom_point(aes(alpha = 0.6), , show.legend = FALSE) +
  geom_smooth(aes(color = Student), method = lm, se = FALSE, show.legend = FALSE) +
  facet_grid(Married ~ Student, labeller = label_both) +
  scale_y_continuous(label = dollar) +
  scale_x_continuous(label = dollar) +
  labs(
    y = "Credit Card Balance"
  )
```



b. Based on your answer to part (a), do you think married and student might be useful predictors, in addition to income for predicting credit card balance? Explain your reasoning.

I believe married & student might be useful predictors in addition to student for predicting credit card balance because f the difference in finances between the two type of predictors. In marriage, credit card accounts are often shared, and can lead to high balances. Students might also rely on credit cards, due to lack of funds in their regular debit account.

c. Credit utilization is defined as the proportion of credit balance to credit limit. Calculate credit utilization for all individuals in the credit data, and use it to recreate the following visualization. Once again, the only aspect of the visualization you do not need to match are the colors, but you should use the same colors from the previous exercise.

```
Credit %>%
  mutate(Credit_Utilization = (Balance/Limit), na.rm = TRUE) %>%
  ggplot(aes(y = Credit_Utilization, x = Income, color = Student, shape = Student, show.lege
  geom_point(aes(alpha = 0.6), , show.legend = FALSE) +
  geom_smooth(aes(color = Student), method = lm, se = FALSE, show.legend = FALSE) +
  facet_grid(Student ~ Married, labeller = label_both) +
  scale_y_continuous(label = percent) +
  scale_x_continuous(label = dollar) +
  labs(
```

```
    y = "Credit Utilization"
  )
```

d. Based on the plot from part (c), how, if at all, are the relationships between income and credit utilization different than the relationships between income and credit balance for individuals with various student and marriage status.

The difference in relationships between credit ussage & credit balance are different because there seems to be a direct relationship between with all the predictors in Credit Balance, however that relationship is either a less significant direct relationship, or a indirect relationship.

Exercise 3

a) Prep - Dataframe Funsies

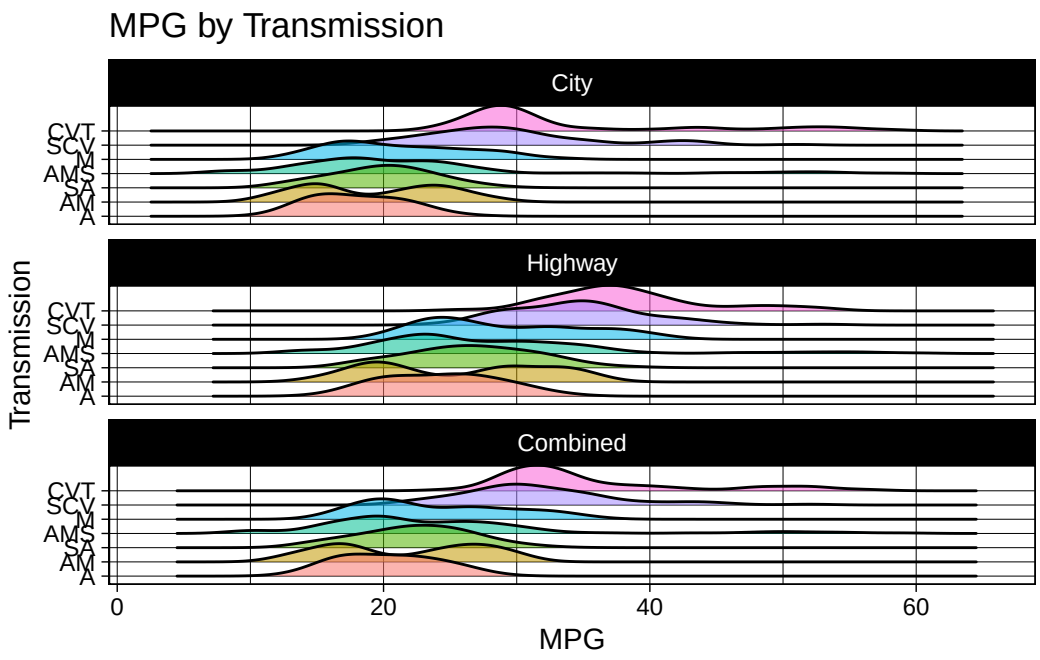
```
epa2021_longer <- epa2021 %>%
  pivot_longer(
    cols = ends_with("_mpg"),
    values_to = "mpg_value",
    names_to = "mpg_type"
    )

epa2021_longer$mpg_type <- as.factor(epa2021_longer$mpg_type)
```

```
epa2021_longer <- epa2021_longer %>%
  mutate(
    mpg_type = fct_relevel(mpg_type, "city_mpg", "hwy_mpg",  "comb_mpg"),
    mpg_type = fct_recode(mpg_type, City = "city_mpg", Highway = "hwy_mpg", Combined = "comb
    transmission = fct_reorder(transmission, mpg_value, .fun = mean)
)
```

b) Visulaize

```
epa2021_longer %>%
  ggplot(aes(y = transmission, x = mpg_value)) +
  geom_density_ridges(aes(fill = transmission, alpha = 0.6), show.legend = FALSE) +
  facet_wrap(~mpg_type, nrow = 3) +
  theme_linedraw() +
  labs(
    title = "MPG by Transmission",
    y = "Transmission",
    x = "MPG"
  )
```



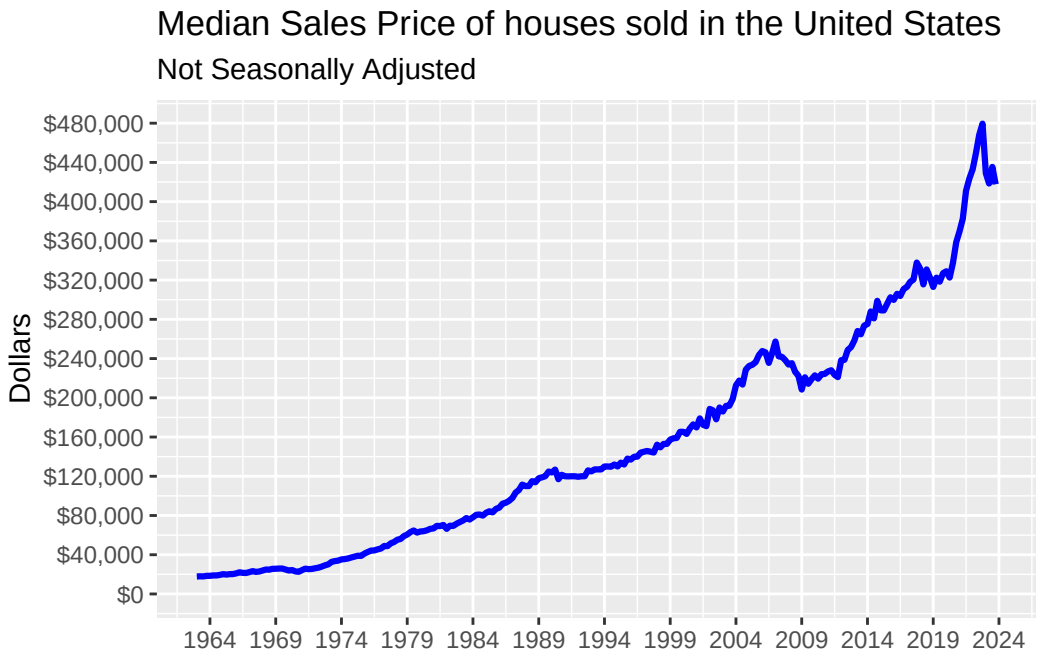
Exercise 4

a). Loading Datasets

```
median_housing <- read_csv("data/median-housing.csv")
recessions <- read_csv("data/recessions.csv")
```

b). Create the following visualization.

```
median_housing %>%
  ggplot(aes(x = date, y = price)) +
  geom_line(color = "blue", size = 1.1) +
  scale_y_continuous(labels = dollar, limits = c(0,480000), breaks = breaks_width(40000)) +
  scale_x_date(date_labels = "%Y", breaks = breaks_width("5 years"))  +
  labs(
    title = "Median Sales Price of houses sold in the United States",
    subtitle = "Not Seasonally Adjusted",
    y = "Dollars",
    x = NULL
    )
```



c). Recessions Truth or Dare

```r
recessions <- recessions %>%
  mutate(
    recessions = ifelse(trough >= min(median_housing$date) &
                          trough <= max(median_housing$date),
                          TRUE, FALSE)
  )
```

d). Another Visulization