Homework 4

This homework is due on Feb. 27, 2025 at 11:00pm. Please submit as a pdf file on Canvas.

In this homework, we will work with the `ufo_sightings` dataset:

```
head(ufo_sightings)
```

```
# A tibble: 6 × 13
  month day    year city                state country shape    duration_seconds
  <chr> <chr> <dbl> <chr>               <chr> <chr>   <chr>              <dbl>
1 10    10     1949 san marcos          TX    us      cylinder            2700
2 10    10     1955 chester (uk/england) <NA>  gb      circle                20
3 10    10     1956 edna                TX    us      circle                20
4 10    10     1960 kaneohe             HI    us      light                900
5 10    10     1961 bristol             TN    us      sphere               300
6 10    10     1965 penarth (uk/wales)   <NA>  gb      circle               180
# ℹ 5 more variables: duration_hours_min <chr>, comments <chr>,
#   year_posted <chr>, latitude <dbl>, longitude <dbl>
```

The main columns we will use are `year` (the year of the sighting), `city` (the city in which the sighting was reported), and `state` (the state in which the sighting was reported).

Problem 1: (4 pts) Since 1940 (inclusive), what are the top 5 cities that have reported the most UFO sightings? Create a new dataframe to answer the question. No plots are necessary.

(Hint: You can use `slice(1:5)` to select the first five rows in a data frame.)

```
# your code here
```

Problem 2: (12 pts)

Using your data frame from Problem 1, make a pie chart of the relative proportions of UFO sightings within the top five cities. Use the manual method of pie chart creation discussed in class (link). Customize the plot so it looks nice. In particular, add labels to each pie slice and remove the separate legend (hint: `guide(fill = "none")`). Also use `scale_fill_manual()` to customize the fill colors and `theme_void()` to remove the grid and axes.

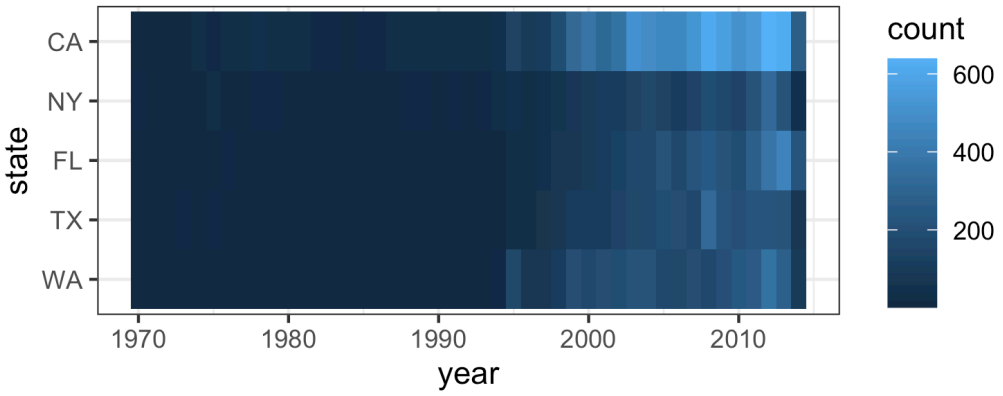
```
# your code here
```

Problem 3: (4 pts)

The following is a plot of the number of UFO sightings per year since 1970 in the top-five states with the most UFO sightings. Modify the plot by adding an appropriate color scale from the `colorspace` package. Then explain in 2-3 sentences why you picked this scale function.

```
ufo_top_five <- ufo_sightings |>
  filter(state %in% c("CA", "WA", "FL", "TX", "NY"), year >= 1970) |>
  count(year, state) |>
  mutate(state = fct_reorder(state, n)) |>
  select(year, state, count = n)

ggplot(ufo_top_five, aes(year, state, fill = count)) +
  geom_tile() +
  theme_bw()
```



Your explanation here.