Homework 5

Enter your name and EID here

Connor Hanna cdh3663

This homework is due on Feb. 28, 2022 at 11:00am. Please submit as a pdf file on Canvas.

Problem 1: (4 pts) We will work with the `ufo_sightings` dataset.

Since 1990, 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.

```
ufo_sightings
```

```
## # A tibble: 70,662 x 13
##    month day    year city  state country shape duration_seconds duration_hours_~
##    <chr> <chr> <dbl> <chr> <chr> <chr>   <chr>            <dbl> <chr>
##  1 10    10     1949 san ~ TX    us      cyli~             2700 45 minutes
##  2 10    10     1955 ches~ <NA>  gb      circ~               20 20 seconds
##  3 10    10     1956 edna  TX    us      circ~               20 1/2 hour
##  4 10    10     1960 kane~ HI    us      light              900 15 minutes
##  5 10    10     1961 bris~ TN    us      sphe~              300 5 minutes
##  6 10    10     1965 pena~ <NA>  gb      circ~              180 about 3 mins
##  7 10    10     1965 norw~ CT    us      disk              1200 20 minutes
##  8 10    10     1966 pell~ AL    us      disk               180 3  minutes
##  9 10    10     1966 live~ FL    us      disk               120 several minutes
## 10 10    10     1968 hawt~ CA    us      circ~              300 5 min.
## # ... with 70,652 more rows, and 4 more variables: comments <chr>,
## #   year_posted <chr>, latitude <dbl>, longitude <dbl>
```

```
#so the observations are for individual observations... let's fix that

city_table <- table(ufo_sightings$city)
city_table <- arrange(as.data.frame(city_table), desc(Freq))
city_table <- head(city_table, 5)
city_table
```

```
##          Var1 Freq
## 1     seattle  524
## 2     phoenix  454
## 3    portland  373
## 4   las vegas  367
## 5 los angeles  352
```

```
#cool, now we have the top five
```

Problem 2: (4 pts)

Next, we will be looking at how the number of UFO sightings has changed for five states since 1940 for both problems 2 and 3. Please follow these steps:

1. Filter the dataset to keep the following five states: TX, CA, NM, FL, NY
2. Keep only the records from 1940 and onwards.
3. Find the number of records for each year and state.
4. Display the new table below your code block

```
# filtering for the specified states
state_list <- c("TX", "CA", "NM", "FL", "NY")
ufo_2 <- filter(ufo_sightings, str_detect(ufo_sightings$state, state_list))
```

```
## Warning in stri_detect_regex(string, pattern, negate = negate, opts_regex =
## opts(pattern)): longer object length is not a multiple of shorter object length
```

```
table(ufo_2$state)
```

```
##
##   CA   FL   NM   NY   TX
## 1798  794  130  579  720
```

```
# keeping only records from 1940 onwards
ufo_3 <- filter(ufo_2, year >= 1940)
table(ufo_3$year)
```

```
##
## 1945 1946 1947 1949 1951 1952 1953 1954 1955 1956 1957 1959 1960 1961 1962 1963
##    1    2    2    1    2    3    1    4    2    2    3    1    2    1    3    3
## 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979
##    4   11   10   13    2    7   11    3    7    6    9   15   15   20   13   19
## 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995
##    8    9    9    6   10   14   10   10    9   15   13   11   16   18   19   57
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##   38   52   80  120  153  130  165  228  196  204  214  230  305  257  212  261
## 2012 2013 2014
##  315  308  119
```

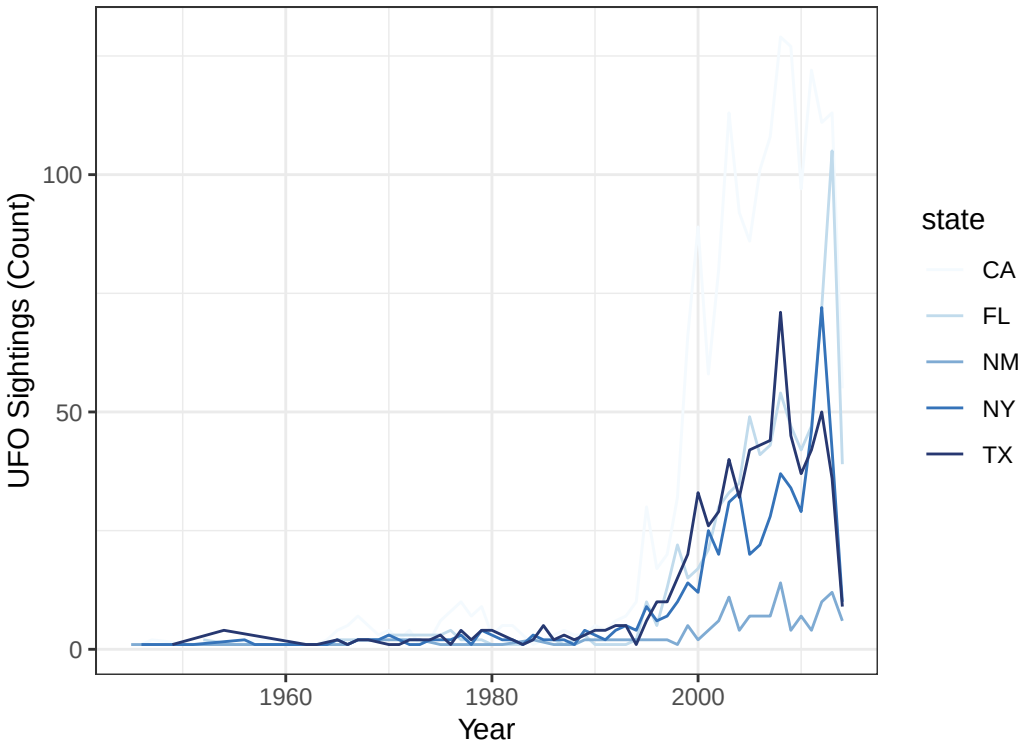
```
# now we need a table grouped by state and year:
table <- as.data.frame(table(ufo_3$year, ufo_3$state))
table <- rename(table, year = Var1, state = Var2)

# the next problem seems to require mapping the result from the table onto the y axis... but those resu
ufo_3 <- merge(ufo_3, table, by = c("year", "state"))
```

Problem 3: (2 pts)

Use the new dataframe you made in Problem 2 and add an appropriate color scale from the `colorspace` package to the plot below.

```
ufo_3 %>% # use the dataframe from Problem 2 here, and set eval = TRUE in the chunk header
 ggplot(aes(x = year, y = Freq, color = state)) +
 geom_line() +
 xlab("Year") +
 ylab("UFO Sightings (Count)") +
 theme_bw() +
 scale_color_discrete_sequential(palette = "Blues")
```


UFO Sightings (Count)
100
50
0
1960
1980
2000
Year
state
CA
FL
NM
NY
TX