```r
# Load data
library(sf)

NC <- read_sf("data/NC_REGION.shp")
```

R Module 7

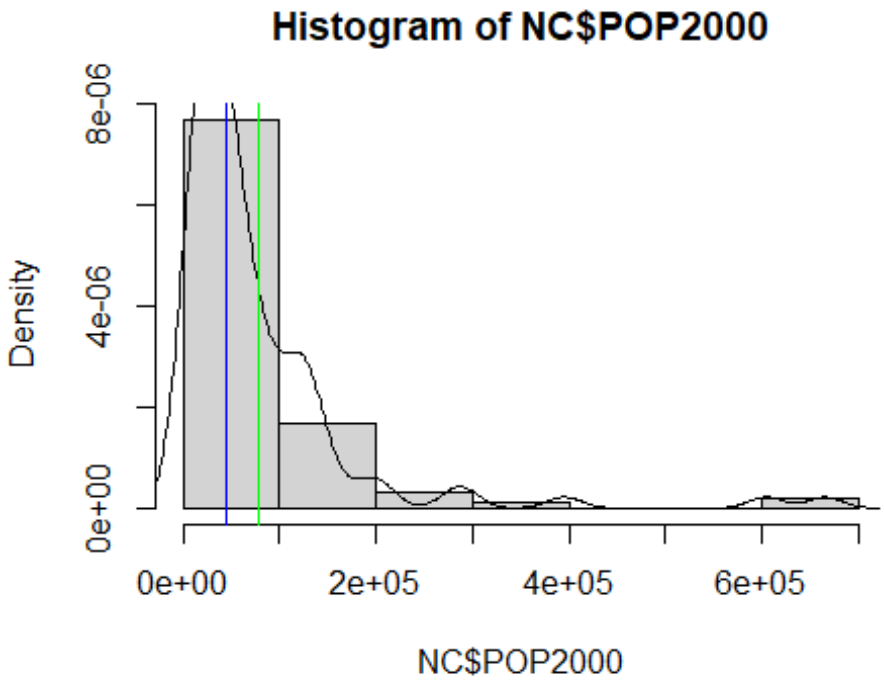
As in previous R Modules, write up this document as an R Markdown report, and export the results to a `.pdf`. Include both your results, your R code, and the answers to the questions.

Question 1:

Create a histogram for the `POP2000` variable. Include the histogram itself, the density, the mean and median lines, and the axis labels and title.

Base R:

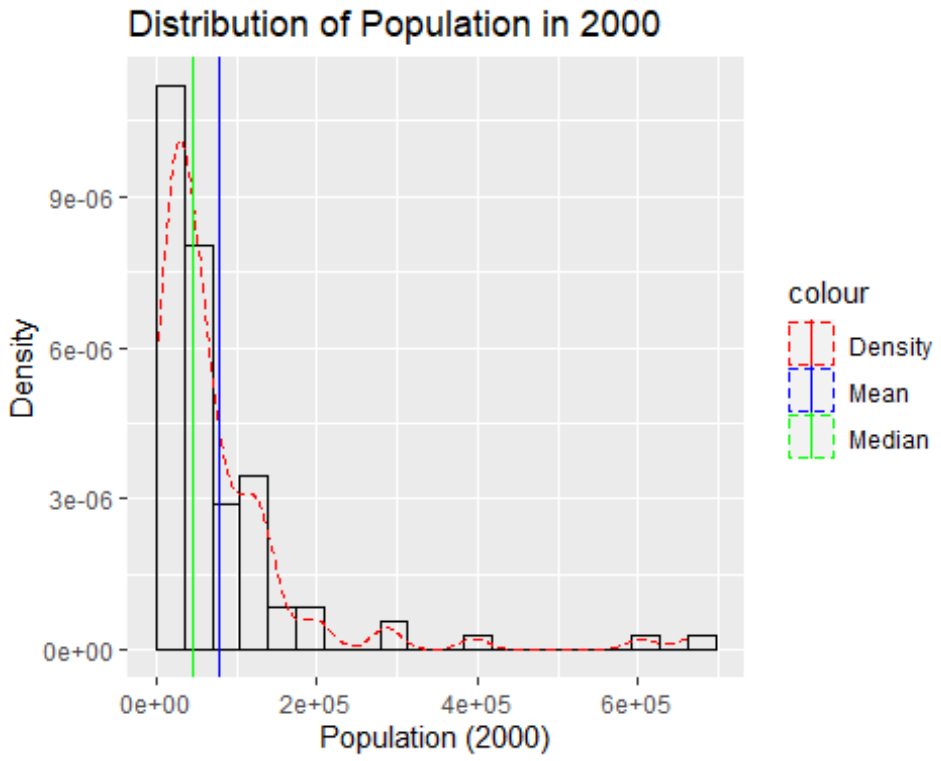
```r
hist(NC$POP2000, probability = T)
lines(density(NC$POP2000))
abline(v = mean(NC$POP2000), col = "green")
abline(v = median(NC$POP2000), col = "blue")
```



ggplot2:

```r
library(ggplot2)
```

```r
# This is pretty tricky to get to with ggplot2, so I don't expect many to get
# this one. By far the most difficult part is getting the legend to display
what
# would be intuitive.


ggplot(NC, aes(x = POP2000)) +
  geom_histogram(aes(y = ..density..),
    fill = NA,
    color = "black",
    boundary = 0,
    bins = 20
  ) +
  geom_density(
    linetype = "dashed",
    aes(color = "Density")
  ) +
  geom_vline(aes(xintercept = mean(POP2000), color = "Mean")) +
  geom_vline(aes(xintercept = median(POP2000), color = "Median")) +
  scale_color_manual(values = c(
    "Density" = "red",
    "Mean" = "blue",
    "Median" = "green"
  )) +
  labs(
    x = "Population (2000)",
    y = "Density",
    title = "Distribution of Population in 2000"
  )
```

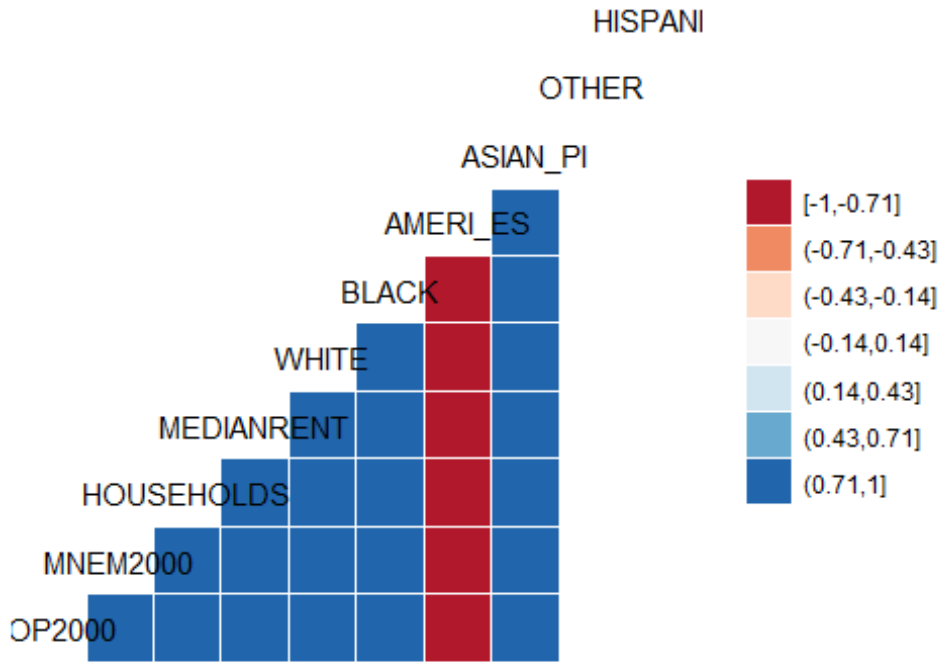


Question 2:

Using the `correlate()` and `fashion()` functions in `corrr`, create a Pearson's *r* correlation matrix of your filtered data. When filtering and selecting columns, choose different/additional columns to compare (don't just use the ones in the Lab!). Provide the matrix and the correlogram (and your R code used to make it!) in your R Markdown report. Identify the strongest and weakest correlation coefficients where $r < 1$.

```
library(dplyr)

NC_filter <- NC %>%
  st_drop_geometry() %>%
  select_if(is.numeric) %>%
  # Students should have a different set of variables, but can use some of
the
  # following.
  dplyr::select(
    c(
      POP2000,
      MNEM2000,
      HOUSEHOLDS,
      MEDIANRENT,
      WHITE,
      BLACK,
      AMERI_ES,
      ASIAN_PI,
```

```
      OTHER,
      HISPANIC
    )
  )

library(corrr)

cor <- NC_filter %>%
  correlate() %>%
  shave()

fashion(cor, decimals = 3)

library(GGally)

cor %>%
  ggcorr(nbreaks = 7, palette = "RdBu")
```



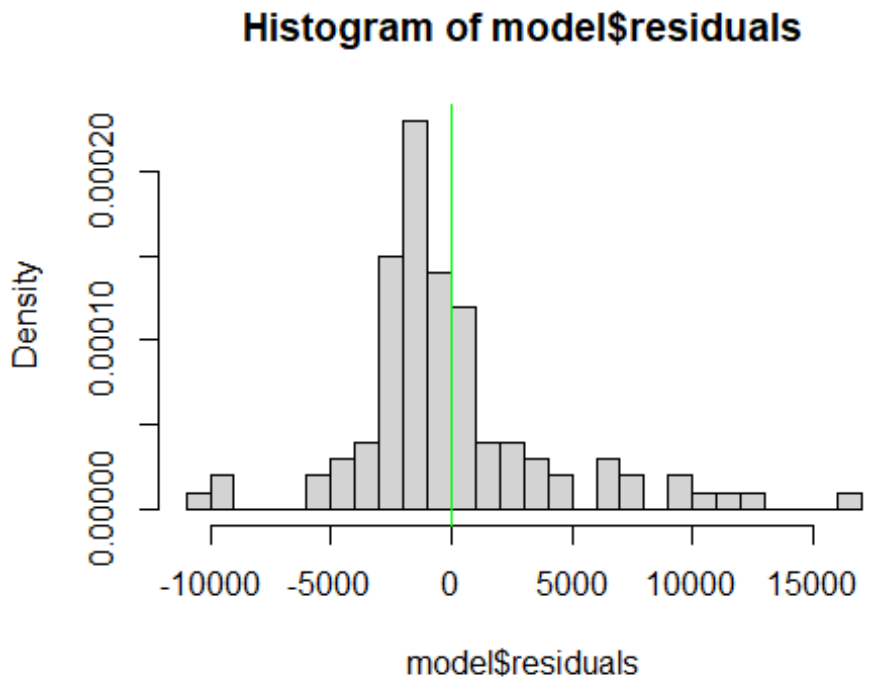
Question 3:

Calculate and report the mean value of the set of residuals from your regression results (in other words, the mean difference between predicted and observed values of our dependent variable). Next, create an absolute value histogram of the residuals from your regression using the 'fd' breaks method. Add the mean value of the residuals to your plot as a vertical line. Include a legend, appropriate axis labels, and a title. As always, provide your R code to do this in your R Markdown report.

```r
model <- lm(MNEM2000 ~ POP2000, data = NC)

m <- mean(model$residuals)
print(paste("Mean residuals = ", m))

## [1] "Mean residuals =  7.8070883091641e-15"
```

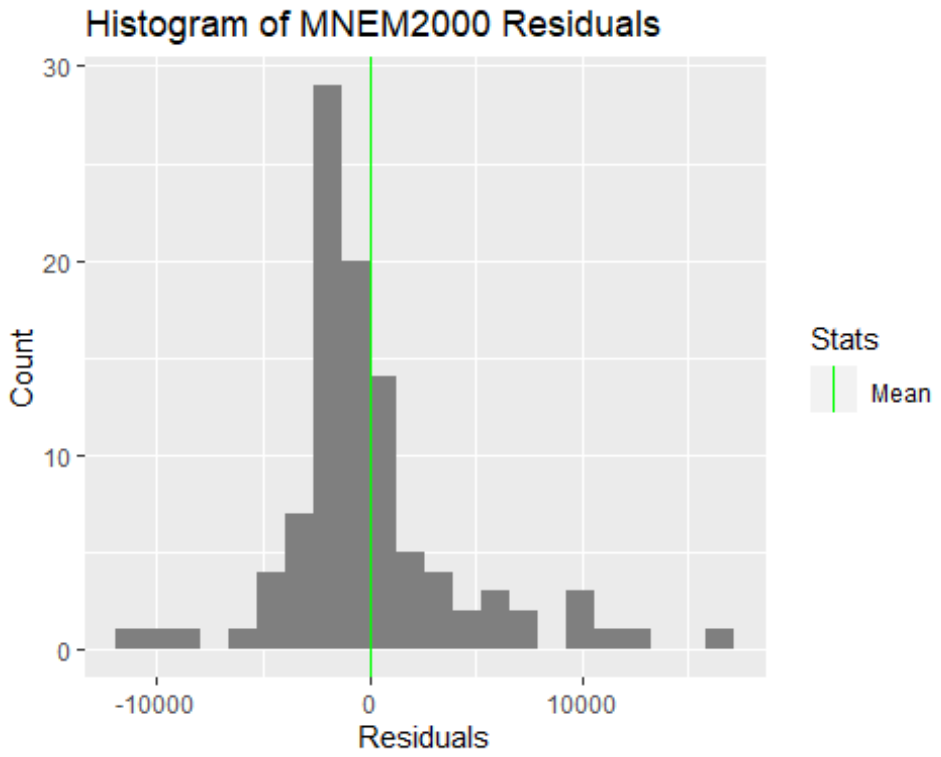
base R:

```r
hist(model$residuals, probability = TRUE, breaks = nclass.FD)
abline(v = mean(model$residuals), col = "green")
```



ggplot2:

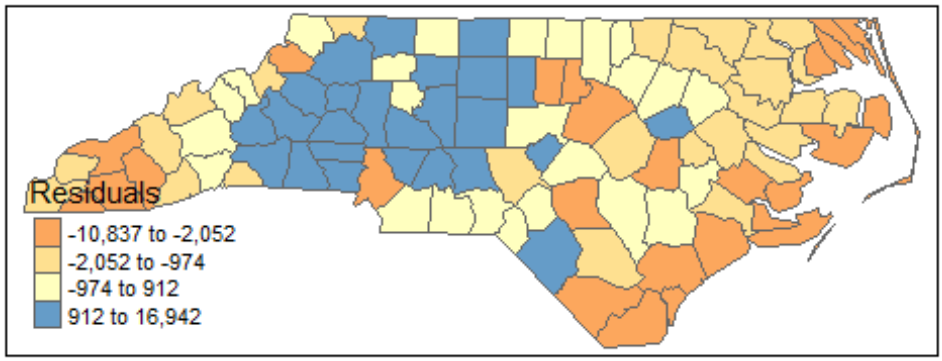
```r
residuals <- data.frame(model$residuals)
fd_bins <- nclass.FD(residuals$model.residuals)
ggplot(residuals, aes(x = model.residuals)) +
  geom_histogram(bins = fd_bins,
                 fill = "gray50",
                 boundary = 0) +
  geom_vline(aes(xintercept = mean(model.residuals), color = "Mean")) +
  scale_color_manual(values = c("Mean" = "green")) +
  labs(
    x = "Residuals",
    y = "Count",
    color = "Stats",
    title = "Histogram of MNEM2000 Residuals"
  )
```



Question 4:

Create a choropleth map of the residuals of the regression with an appropriate legend and title. You may explore the spatial patterning fo residuals through altering your break methods and number of categories, but for your final map, use 4 classes and a quartiles classification theme. Why? This utilizes the median and 1st/3rd quartiles as breakpoints. In other words, our map will be connected to a measure of central tendency of the residuals, which can aid interpretation. Explain in writing if there is visual evidence for spatial dependence in the map. Provide the map and your R code.

```
library(tmap)
NC$Residuals <- model$residuals

tm_shape(NC) +
  tm_polygons(
    col = "Residuals",
    style = "quantile",
    n = 4,
    palette = "RdYlBu"
  )
```



There seems to be a lot of high residuals in the central part of the state, and low residuals on the coastal plains and western tip regions.