# Lab6_Template

Stutie Banerjie

## Setup

```
# Load packages needed
library(tidyverse)
library(openintro)
library(hexbin)
```

```
# Create condensed dataframe loans
loans <- loans_full_schema %>%
  select(loan_amount, interest_rate, term, grade,
         state, annual_income, homeownership, debt_to_income)
glimpse(loans)
```

```
## Rows: 10,000
## Columns: 8
## $ loan_amount   <int> 28000, 5000, 2000, 21600, 23000, 5000, 24000, 20000, 20…
## $ interest_rate <dbl> 14.07, 12.61, 17.09, 6.72, 14.07, 6.72, 13.59, 11.99, 1…
## $ term          <dbl> 60, 36, 36, 36, 36, 36, 60, 60, 36, 36, 60, 60, 36, 60,…
## $ grade         <fct> C, C, D, A, C, A, C, B, C, A, C, B, C, B, D, D, D, F, E…
## $ state         <fct> NJ, HI, WI, PA, CA, KY, MI, AZ, NV, IL, IL, FL, SC, CO,…
## $ annual_income <dbl> 90000, 40000, 40000, 30000, 35000, 34000, 35000, 110000…
## $ homeownership <fct> MORTGAGE, RENT, RENT, RENT, RENT, OWN, MORTGAGE, MORTGA…
## $ debt_to_income <dbl> 18.01, 5.04, 21.15, 10.16, 57.96, 6.46, 23.66, 16.19, 3…
```

Using the `loans` dataframe create a histogram for `interest_rate`. The histogram should have an appropriate amount of bins. Be sure to provide an apprppriate title and axis labels.

```
ggplot(loans, aes(x = interest_rate)) +
  geom_histogram(binwidth = 1.3) +
  labs(
    x = "Interest Rate (%)",
    y = "Frequency",
    title = "Frequency of Interest Rates)"
  )
```
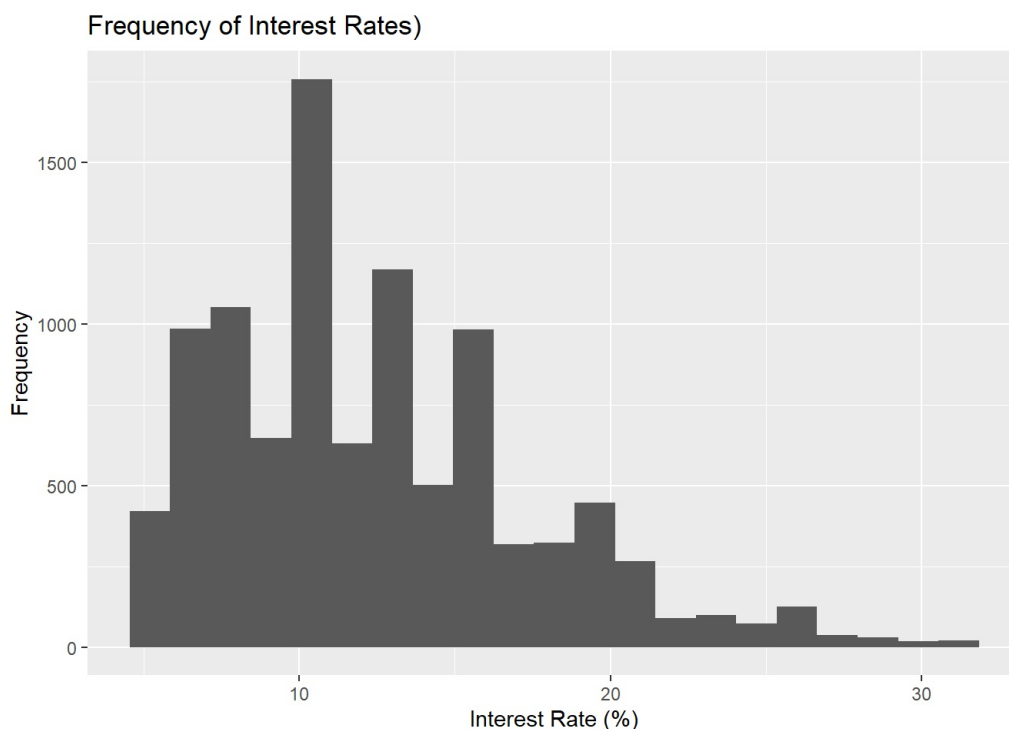


Describe the shape of the distribution of `interest_rate`.

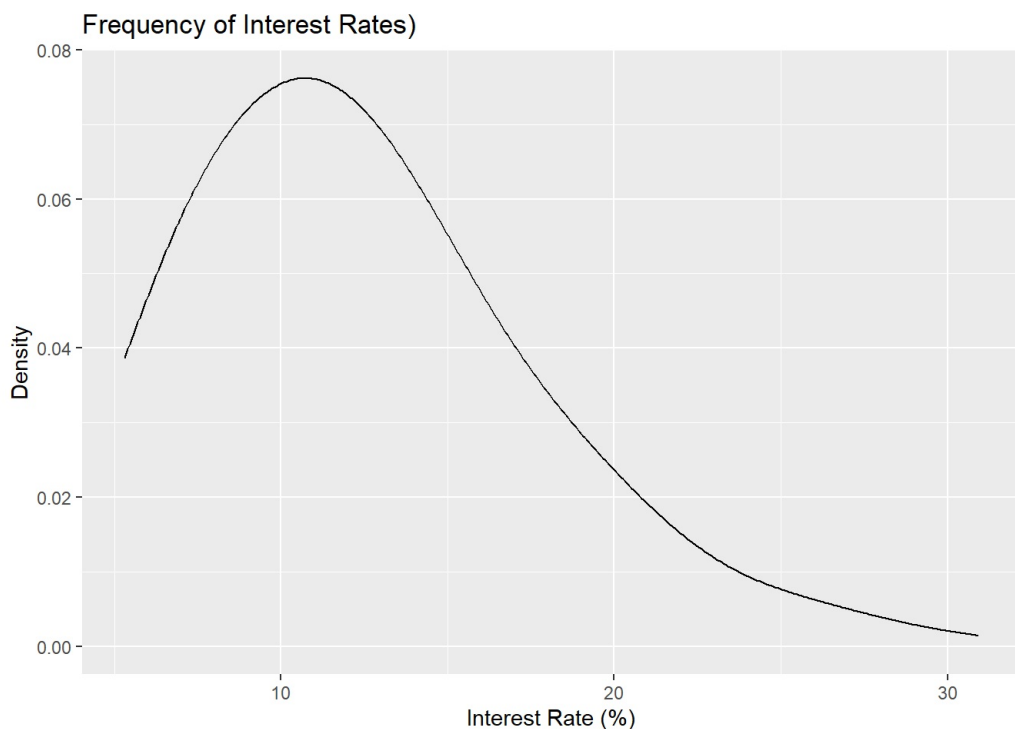The distribution of Interest Rate is skewed right.

Using the `loans` dataframe, create a facet histogram of `interest_rate` broken down by `grade`.

```
ggplot(loans, aes(x = interest_rate)) +
  geom_histogram(binwidth = 2) +
  labs(
    x = "Interest Rate (%)",
    y = "Frequency",
    title = "Frequency of Interest Rates)"
  )+
  facet_wrap(~ grade, nrow=2)
```



Frequency of Interest Rates)
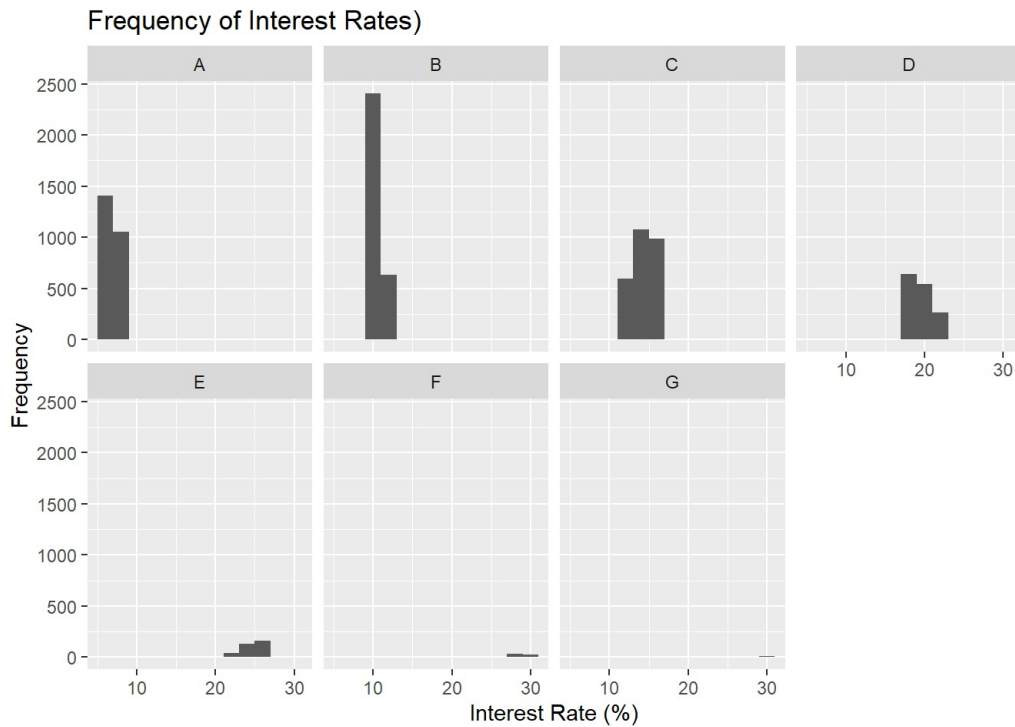
Using the `loans` dataframe, create a density plot of `interest_rate` separated by `grade`. Use an appropriate adjustment value. Use appropriate titles and labels.

```
ggplot(loans, aes(x = interest_rate)) +
  geom_density(adjust=4)+
  labs(
    x = "Interest Rate (%)",
    y = "Density",
    title = "Frequency of Interest Rates)"
  )
```



Frequency of Interest Rates)

Do you notice anything unusual about this graph? Explain.

It's incredibly skewed right, similar to the previously developed histogram. It's interesting to see how the histogram and the density graph mimics each other's shapes despite not being the exact same.
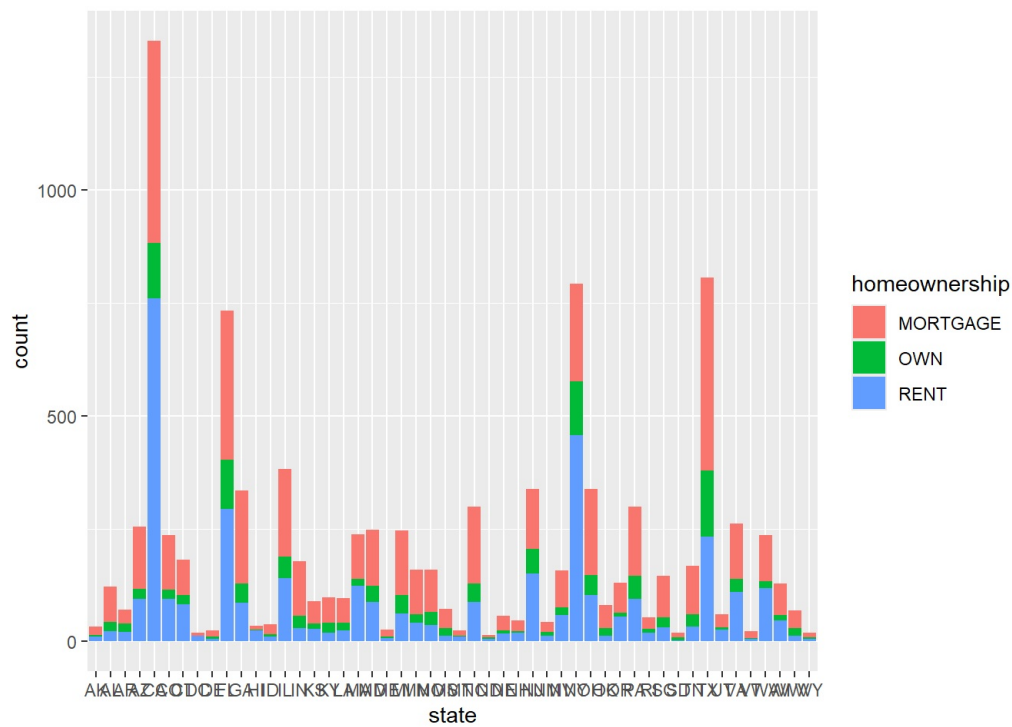
Using the loans dataframe, create boxplots of `interest_rate` sepearted by `grade`, coloring each box a different color based on `grade`. Use appropriate labels and titles.

```
ggplot(loans, aes(x = interest_rate,
                  y = grade, fill=grade)) +
  geom_boxplot() +
  labs(
    x = "Interest Rate (%)",
    y = NULL,
    title = "Distribution of Interest Rates",
    subtitle = "by Grade"
  )
```
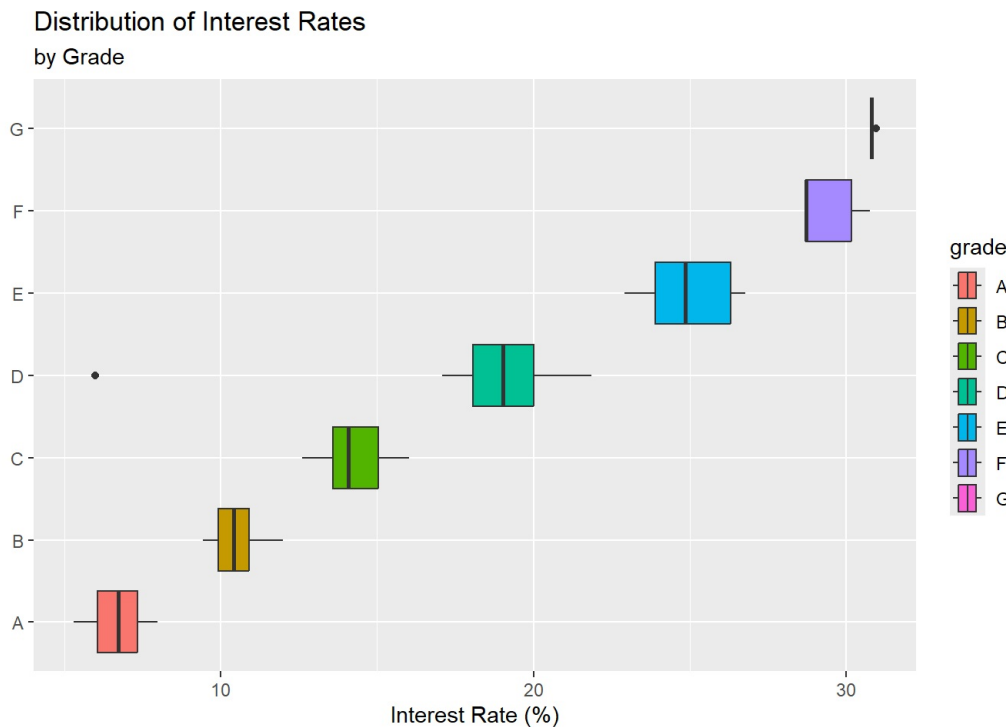


Using the loans dataframe and the idea that you can filter within the ggplot function, create a hexplot of `loan_amount` vs. `interest_rate` for grade (Hint: use the `fill` argument) values A thru F (exclude grade G). Please provide appropriate title(s) and labels.

```
ggplot(loans %>% filter(grade %in% c("A", "B", "C", "D", "E", "F")),
       aes(x = loan_amount, y = interest_rate, fill = grade))+
  geom_hex() +
  labs(title = "Hexplot of Loan Amount vs. Interest Rate",
  x = "Loan Amount ($)",
  y = "Interest Rate (%)",
  fill = "Grade")
```

Hexplot of Loan Amount vs. Interest Rate

Using the `loans` dataframe find the min, med, max, IQR, and num. of observations for `interest_rate` by `grade`. Label your statistics appropriately.

```
loans %>%
  group_by(grade) %>%
  summarise(mean_intRate = mean(interest_rate),
            sd_intRate = sd(interest_rate),
            min_intRate = min(interest_rate),
            max_intRate = max(interest_rate)
            )
```

```
## # A tibble: 7 × 5
##    grade mean_intRate sd_intRate min_intRate max_intRate
##    <fct>        <dbl>      <dbl>       <dbl>       <dbl>
## 1 A             6.74      0.903        5.31        7.97
## 2 B            10.5       0.895        9.43       12.0
## 3 C            14.2       1.18        12.6        16.0
## 4 D            19.1       1.58         6          21.8
## 5 E            25.1       1.23        22.9        26.8
## 6 F            29.4       0.790       28.7        30.8
## 7 G            30.8       0.0433      30.8        30.9
```

Using the `loans` dataframe, create a segmented bar chart of `state` segmented by `homeownership`. The bars should be horizontal (not vertical) and the length of the bars should be the count for each state. The graph should have appropriate titles, and labels.

```
ggplot(loans, aes(x = state,
                  fill = homeownership)) +
  geom_bar()
```

Using the `loans` dataframe, create a frequency table of `grade`.

```
loans %>%
  count(grade)
```

```
## # A tibble: 7 × 2
##   grade     n
##   <fct> <int>
## 1 A      2459
## 2 B      3037
## 3 C      2653
## 4 D      1446
## 5 E       335
## 6 F        58
## 7 G        12
```