Homework 4

Miles Tweed

1/27/2021

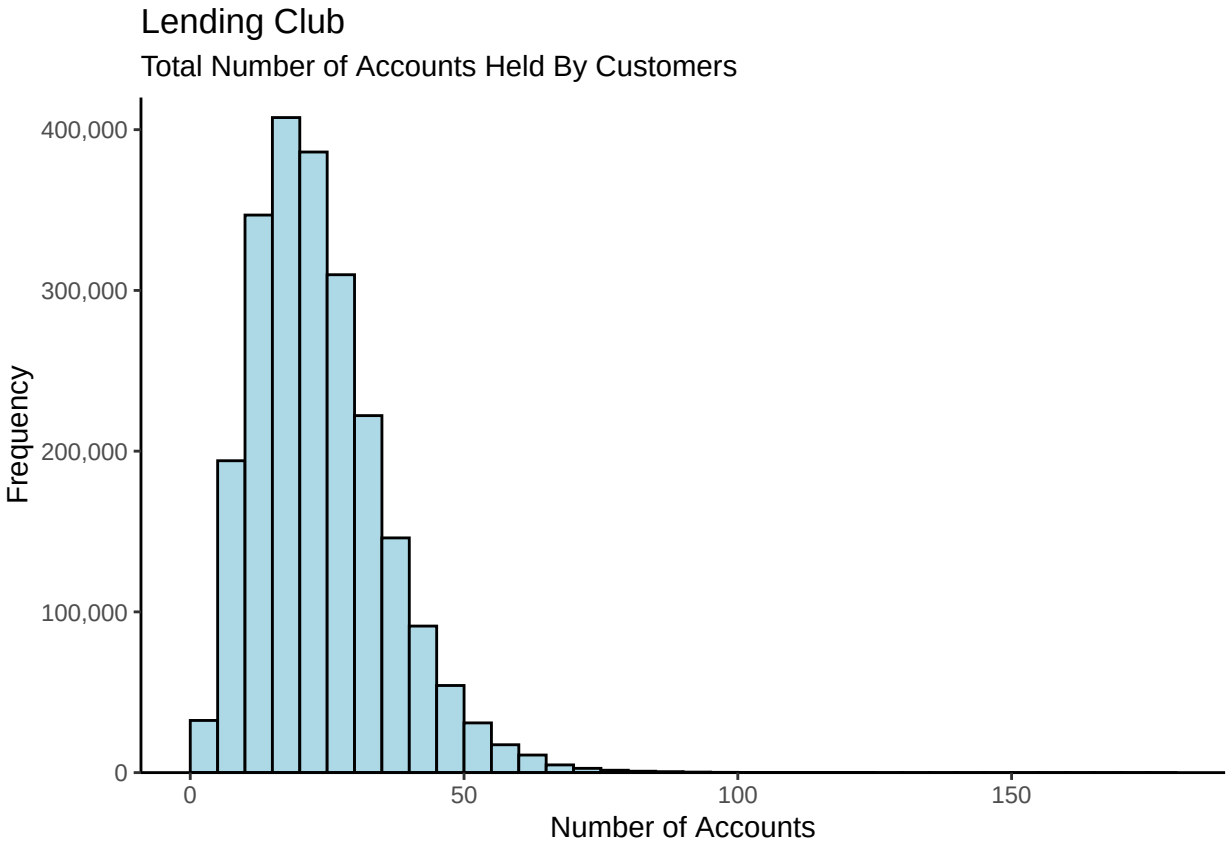
Numerical Sumamry

```
# Total Number of Accounts
lending %>% select(total_acc) %>% drop_na() %>% summary()
```

```
##    total_acc
##  Min.   :  1.00
##  1st Qu.: 15.00
##  Median : 22.00
##  Mean   : 24.16
##  3rd Qu.: 31.00
##  Max.   :176.00
```

Histogram

```
lending %>% select(total_acc) %>%  drop_na() %>%
  ggplot(aes(x = total_acc, y = format(..count.., scientific = FALSE))) +
  geom_histogram(binwidth = 5, boundary = 0, closed = 'right', fill = 'lightblue', color = 'black') +
  scale_y_continuous(limits = c(0,4.2e5), expand = c(0,0), labels = comma) +
  xlab('Number of Accounts') + ylab('Frequency') +
  labs(title = 'Lending Club', subtitle = 'Total Number of Accounts Held By Customers') +
  theme_classic()
```



Distribution

The distribution is unimodal, right skewed, and centered around 22 accounts. The majority of customers have between 0 and 50 accounts while some customers have have as many as 176 accounts. These customers could be considered outliers.