MIS431 Spring 2022 - Homework - 1

Submission Date - 03/13/2022

In this homework assignment we will be focusing on data analysis and visualization with the **tidyverse**.

To complete this assignment, students must use the template available in RStudio Cloud.

Load Packages and Data

The R code chunk below will load the tidyverse and tidymodels packages as well as an auto_claims data set.

```
library(tidyverse)

auto_claims <- read_rds('C:/RDataFiles/auto_claims.rds')
```

Problem 1

Write one expression using dplyr functions and the %>% operator to create the summary table below, and arrange it based on the customer state.

This table contains the number of claims, maximum claim amount (total_claim_amount variable), and minimum customer lifetime value by customer_state and months_since_last_claim binned into 12 month categories.

Hint: You will need to create the month_category variable using cut_width() before you calculate the summaries by groups.

```
## `summarise()` has grouped output by 'customer_state'. You can override using the `.groups` argument.
```

```
## # A tibble: 15 x 5
## # Groups:   customer_state [5]
##    customer_state month_category n_claims max_claim_amount min_clv
##    <fct>          <fct>             <int>            <dbl>   <dbl>
##  1 Arizona        [0,12]              553            3893.   2227.
##  2 Arizona        (12,24]             365            3295.   2255.
##  3 Arizona        (24,36]             263            2550.   2309.
##  4 Nevada         [0,12]              260            2979.   2227.
##  5 Nevada         (12,24]             202            2337.   2329.
##  6 Nevada         (24,36]             139            3552.   2523.
##  7 California     [0,12]              977            3453.   2120.
##  8 California     (12,24]             712            2767.   2301.
##  9 California     (24,36]             461            2361.   2392.
## 10 Oregon         [0,12]              824            2836    2227.
## 11 Oregon         (12,24]             569            2973.   2280.
## 12 Oregon         (24,36]             370            3052    2174.
```

```
## 13 Washington     [0,12]              241           2966.   2227.
## 14 Washington     (12,24]             195           2429.   2301.
## 15 Washington     (24,36]             118           2685.   2472.
```

Problem 2

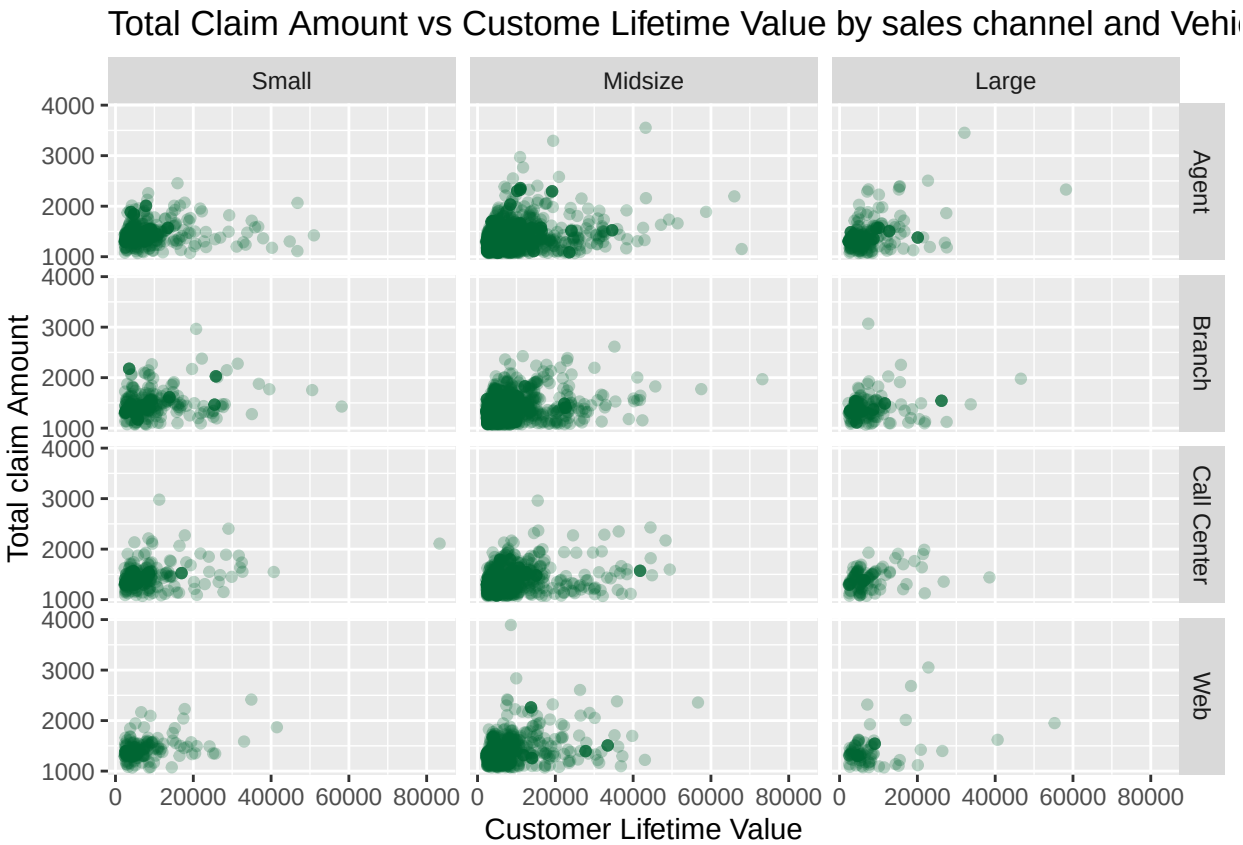
Write one expression using dplyr functions and the %>% operator to create the summary table below.

This table contains the average, median, and maximum values of total_claim_amount by vehicle_class.

```
## # A tibble: 6 x 4
##   vehicle_class ave_amount median_amount max_amount
##   <fct>              <dbl>         <dbl>      <dbl>
## 1 Two-Door Car       1343.         1334.      2258.
## 2 Four-Door Car      1338.         1326.      2266.
## 3 Sports Car         1497.         1523.      2404
## 4 SUV                1511.         1528       2429.
## 5 Luxury Car         1974.         1981.      3295.
## 6 Luxury SUV         2024.         2008       3893.
```

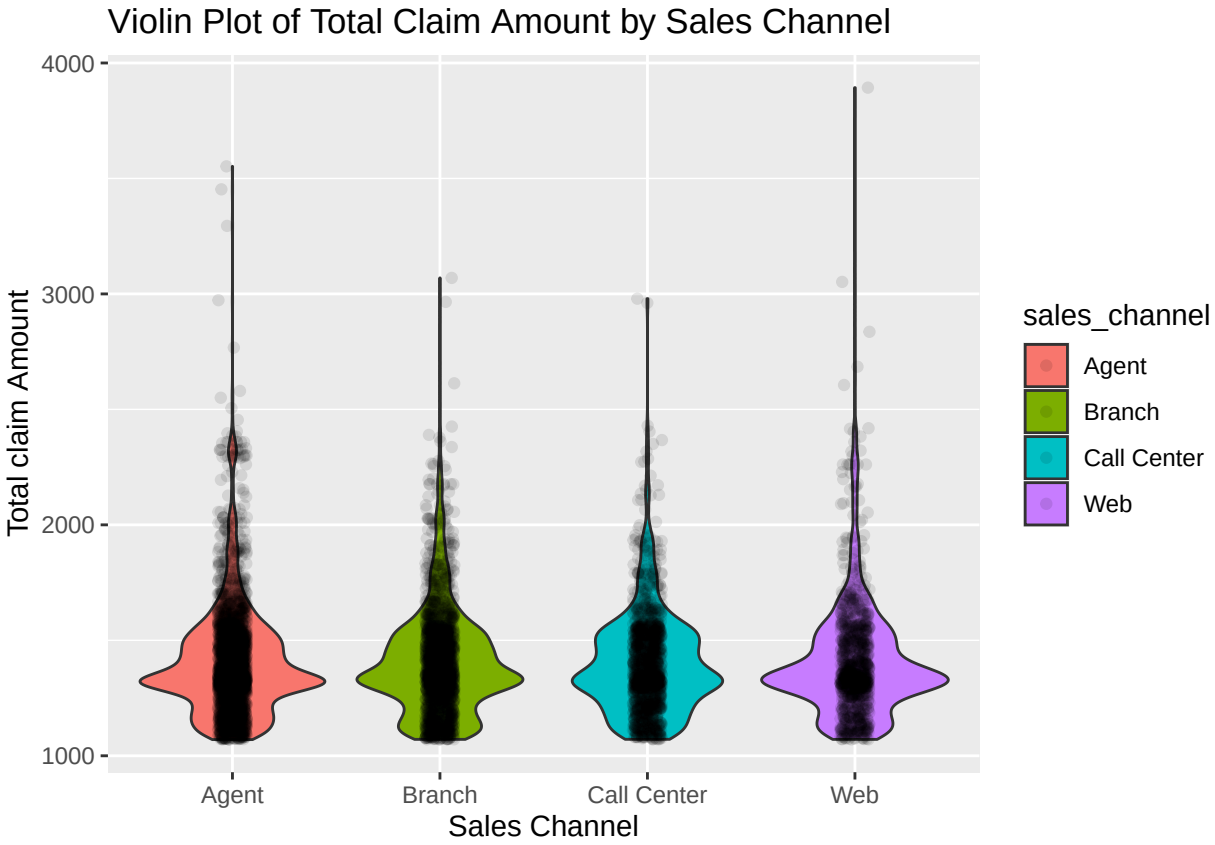
Problem 3

Use ggplot to create the data visualization below. Please use alpha = 0.25 and color = '#006633' within your geom function to match the points below.



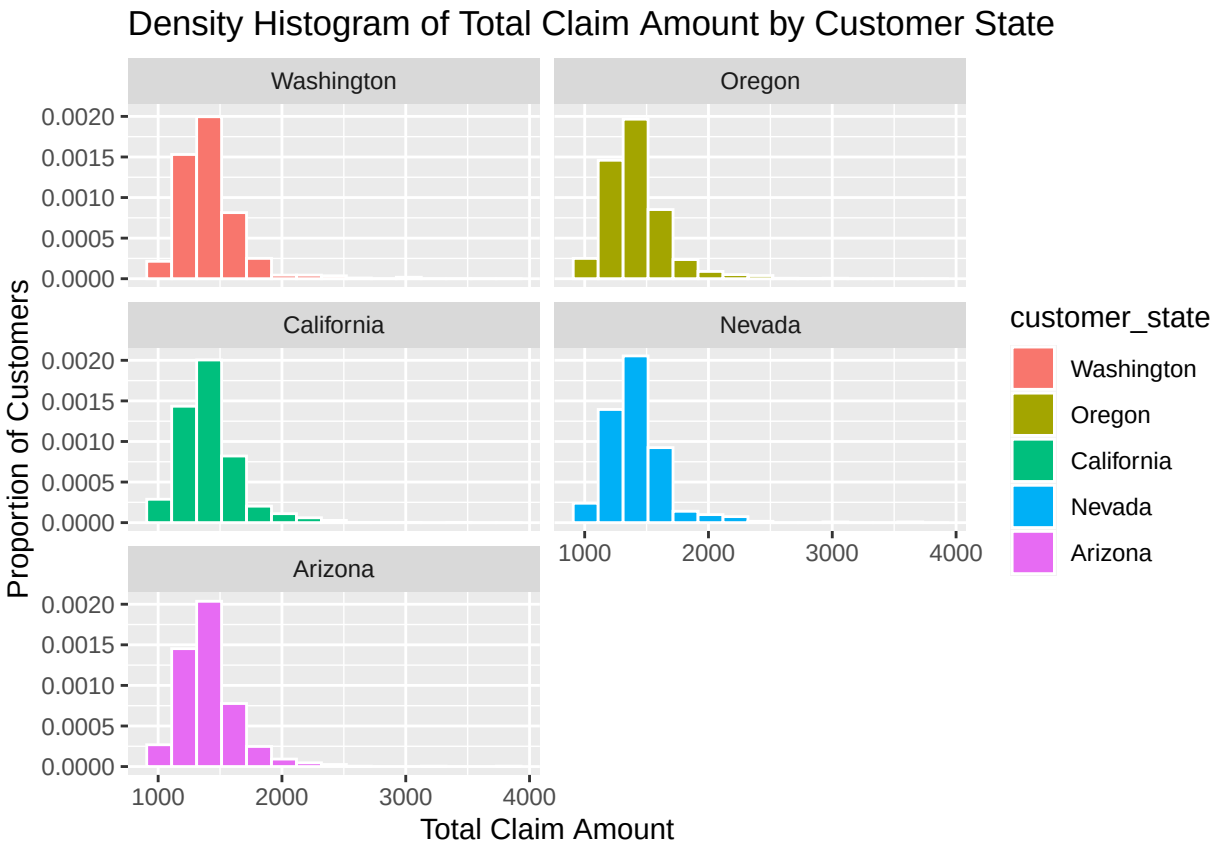
Problem 4

Use ggplot to create the data visualization below.Width value is 0.07 and the aplha value is 0.1



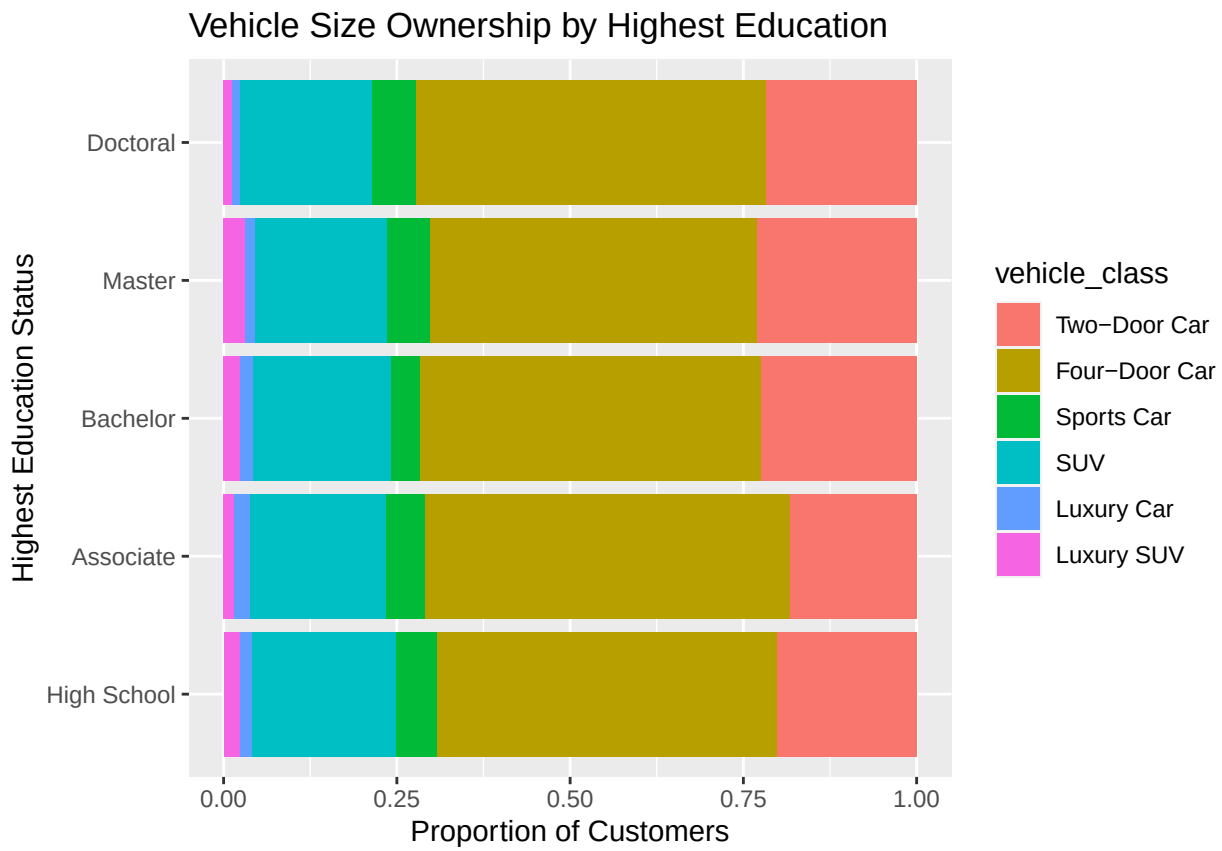
Problem 5

Use ggplot to create the data visualization below. The number of bins in the histogram should be set to 15.



Problem 6

Use ggplot to create the data visualization below.



— End of Homework 1 —