# ST306_LA4_S_18_837

S/18/837

2023-09-26

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.1
```

```
## Warning: package 'readr' was built under R version 4.3.1
```

```
## Warning: package 'lubridate' was built under R version 4.3.1
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(tinytex)
library(ggplot2)
```

```
col_types <- "dccffi"
travel_data <- read_csv("../ST306_LA1_S_18_837/data/travel_data.csv"
                        , col_types = col_types)
```

```
view(travel_data)
```

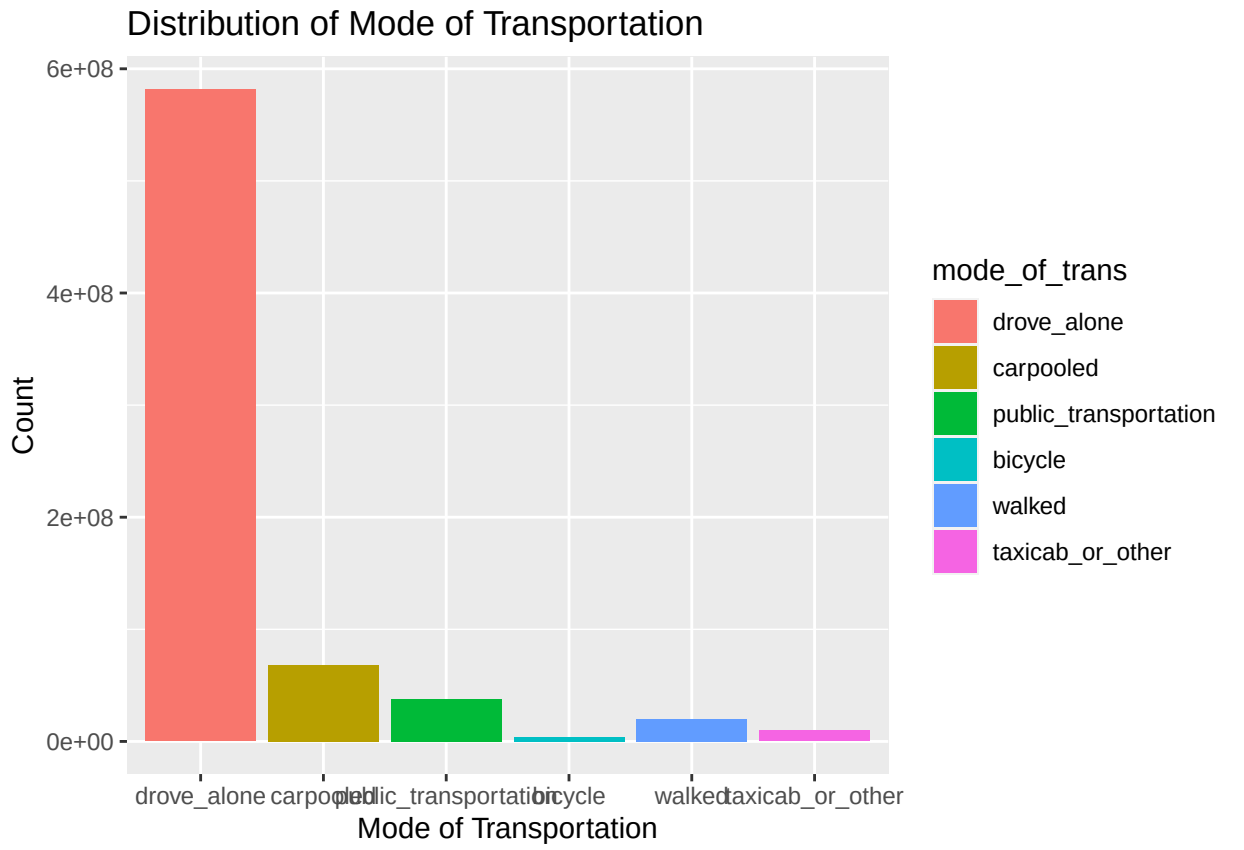
Main Question : What is the most common mode of transportation used to travel from home to work?

```
x1 <- travel_data%>%
  group_by(mode_of_trans)%>%
  summarise(Count = sum(count))%>%
  select(mode_of_trans, Count)%>%
  arrange(desc(Count))
x1
```

```
## # A tibble: 6 x 2
##   mode_of_trans            Count
```

```
##   <fct>                    <int>
## 1 drove_alone            581559564
## 2 carpooled               68235341
## 3 public_transportation   37797191
## 4 walked                  20184995
## 5 taxicab_or_other         9757664
## 6 bicycle                  4130967
```

```
x1 %>%
  ggplot(aes(x=mode_of_trans, y=Count, fill = mode_of_trans))+geom_bar(stat = "identity")+
  labs(title = "Distribution of Mode of Transportation",x = "Mode of Transportation",
       y = "Count")
```




Conclusion : The most common mode of transportation in the travel_data data set is “drove_alone”.

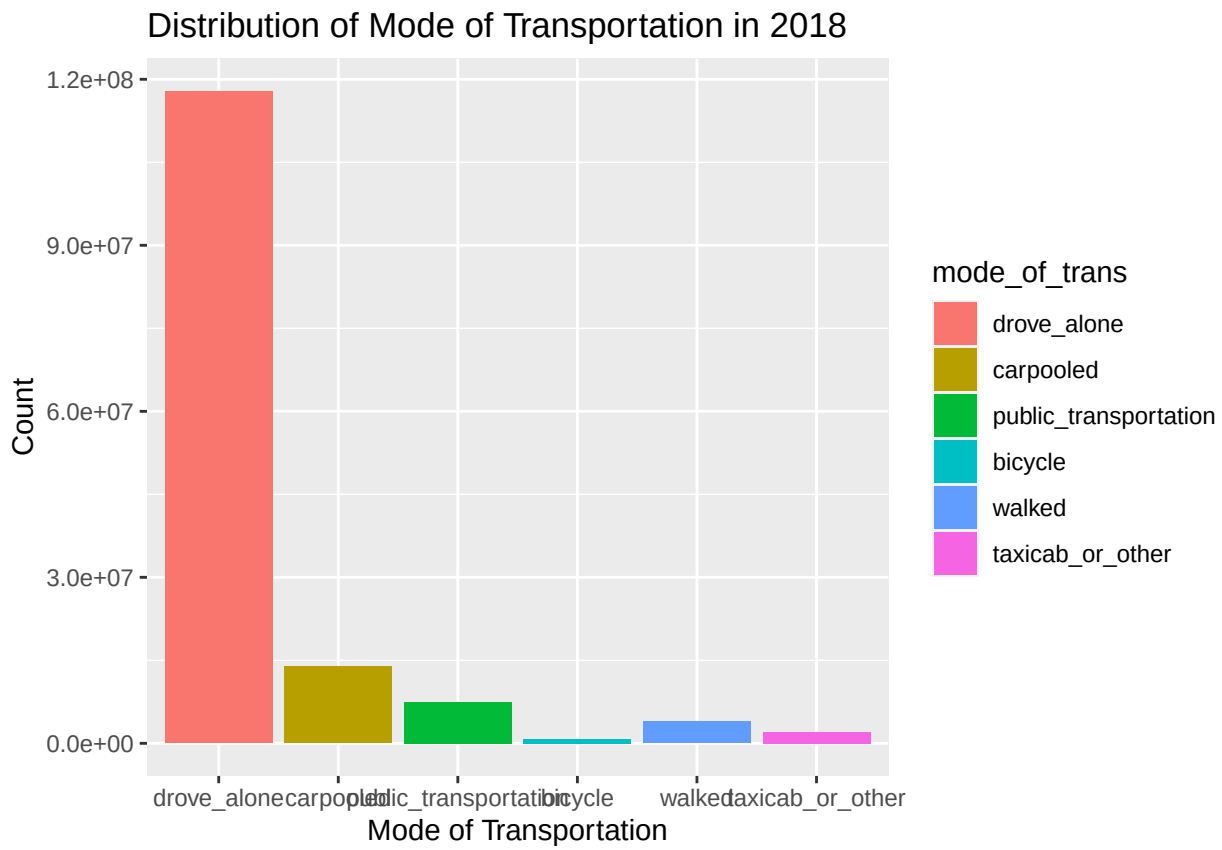
Sub Question 1: Q1)What is the most common mode of transportation used to travel from home to work in 2018?

```
Q1 <- travel_data %>%
  filter(year == 2018) %>%
  group_by(mode_of_trans) %>%
  summarise(Count = sum(count)) %>%
  select(mode_of_trans, Count)
Q1
```

```
## # A tibble: 6 x 2
```

```
##   mode_of_trans             Count
##   <fct>                     <int>
## 1 drove_alone           117818278
## 2 carpooled              13882957
## 3 public_transportation   7484308
## 4 bicycle                  805026
## 5 walked                  3975811
## 6 taxicab_or_other        2036152
```

```
Q1 %>%
  ggplot(aes(x=mode_of_trans, y=Count,fill = mode_of_trans))+geom_bar(stat = "identity") +
  labs(title = "Distribution of Mode of Transportation in 2018",x = "Mode of Transportation",
       y = "Count")
```

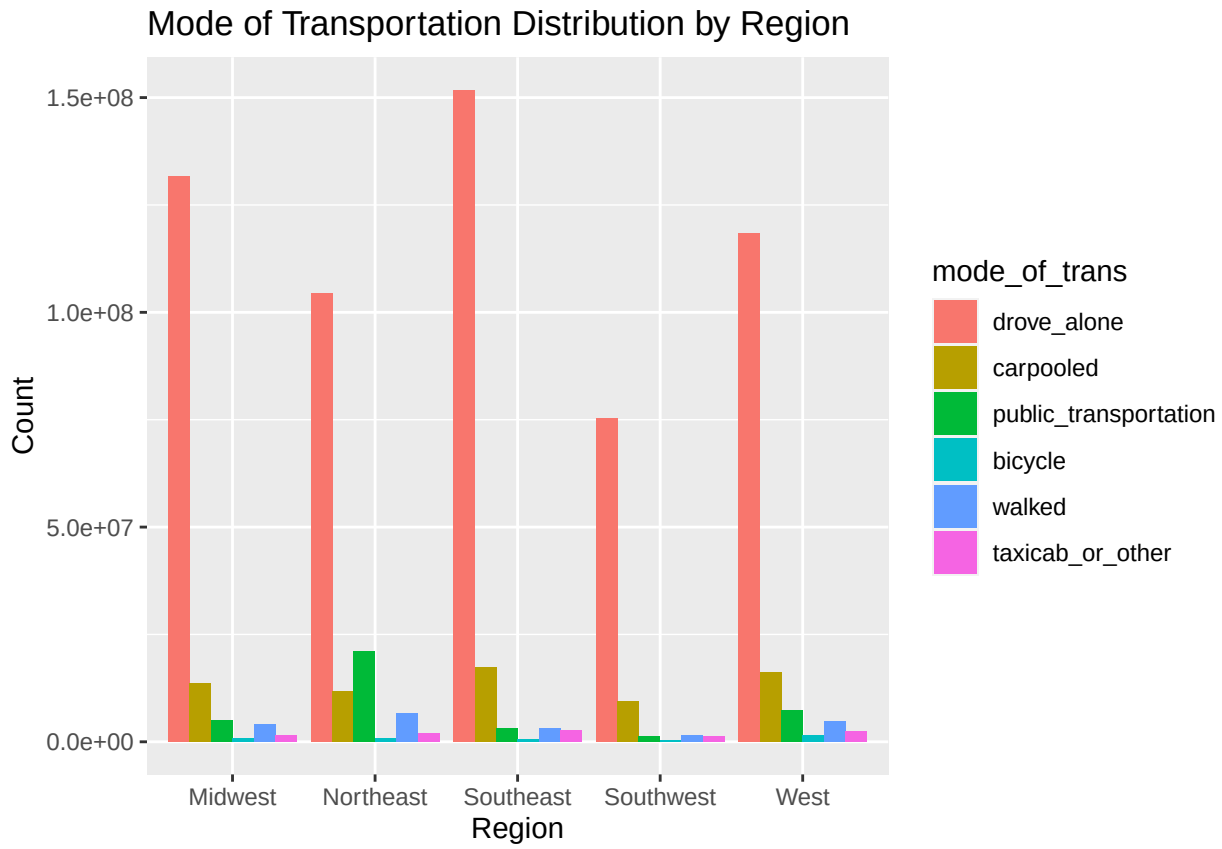



Conclusion : The most common mode of transportation in 2018 in the travel_data data set is "drove_alone".

Sub Question 2: Q2) How does the mode of transportation vary by region?

```
travel_data %>%
  group_by(region, mode_of_trans) %>%
  summarise(count = sum(count)) %>%
  ggplot(aes(x = region, y = count, fill = mode_of_trans)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Mode of Transportation Distribution by Region",
       x = "Region",
       y = "Count")
```

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

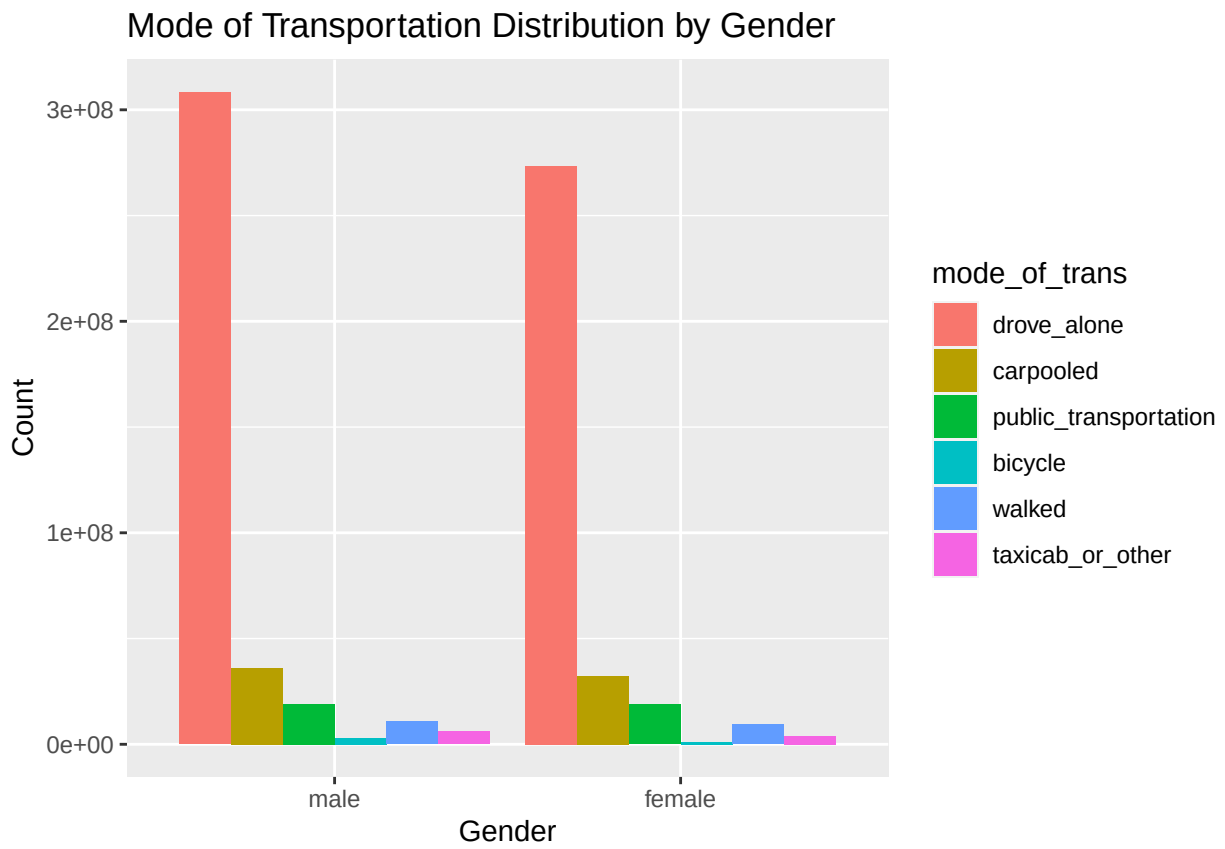



Conclusion : In all regions "drove alone" is the more commonly used method to travel and "bicycle" has the lowest rate.

Sub Question 3: Q3)How does the mode of transportation vary by gender?

```
travel_data%>%
  group_by(gender, mode_of_trans) %>%
  summarise(count = sum(count)) %>%
  ggplot(aes(x = gender, y = count, fill = mode_of_trans)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Mode of Transportation Distribution by Gender",
       x = "Gender",
       y = "Count")
```

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



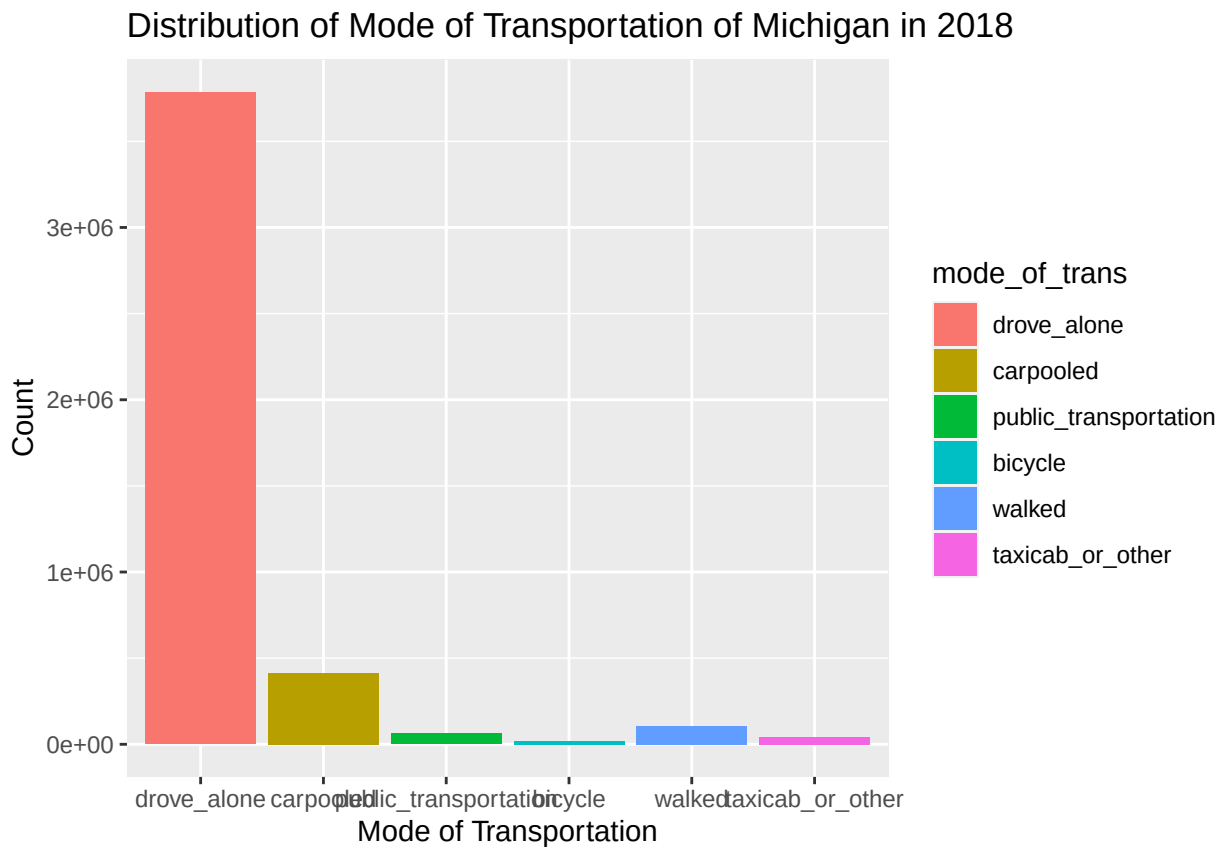

Conclusion : Generally, "drove_alone" is the most common mode for both genders, but the proportions differ.

Sub Question 4: Q4) What is the most common mode of transport in Michigan in 2018?

```
travel_data %>%
  filter(year == 2018 & state == "Michigan")%>%
  group_by(mode_of_trans)%>%
  summarise(Count = sum(count))
```

```
## # A tibble: 6 x 2
##   mode_of_trans           Count
##   <fct>                   <int>
## 1 drove_alone           3784352
## 2 carpooled              412152
## 3 public_transportation   62919
## 4 bicycle                 19400
## 5 walked                 104812
## 6 taxicab_or_other        42216
```

```
travel_data %>%
  filter(year == 2018 & state == "Michigan")%>%
  group_by(mode_of_trans)%>%
  summarise(Count = sum(count))%>%
  ggplot(aes(x=mode_of_trans, y=Count, fill = mode_of_trans))+geom_bar(stat = "identity") +
    labs(title = "Distribution of Mode of Transportation of Michigan in 2018",x = "Mode of Transportati
       y = "Count")
```

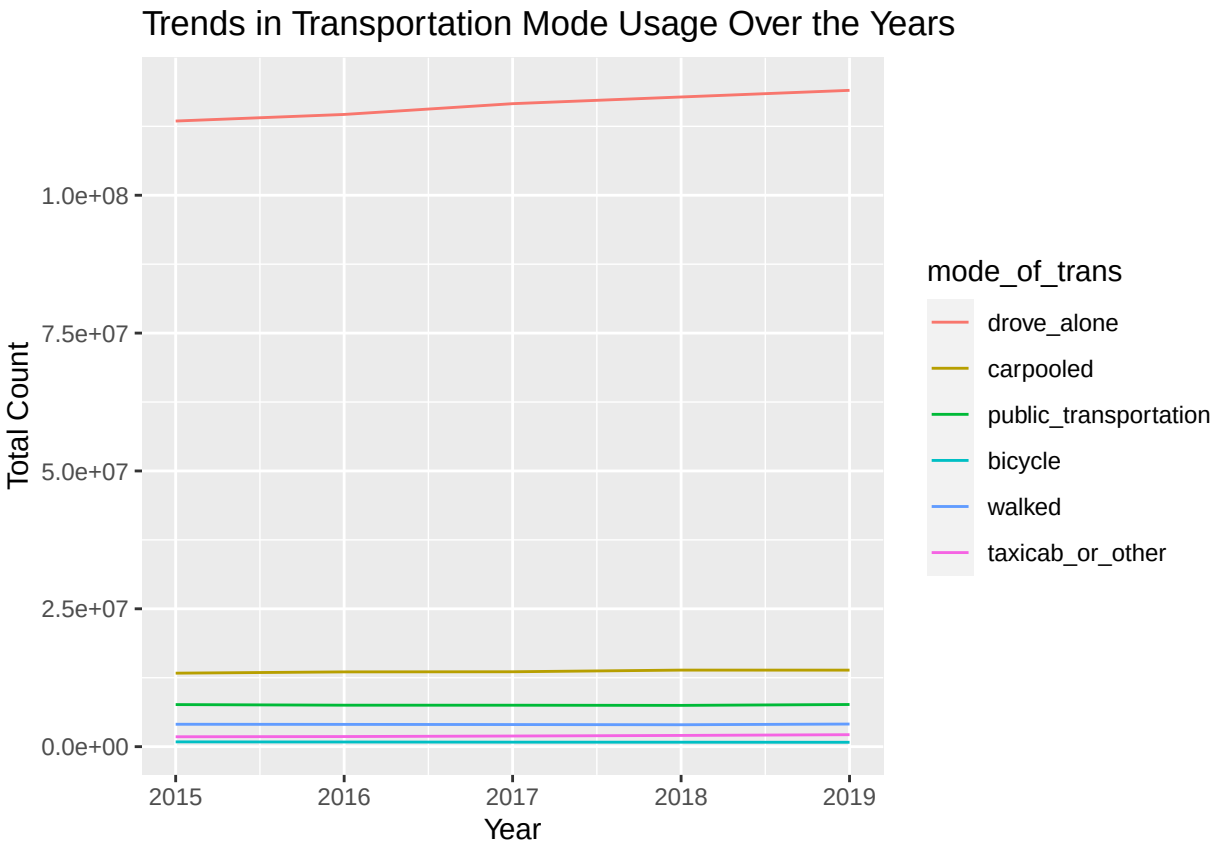



Conclusion : The more commonly transportation mode in Michigan in 2018 is drove alone followed by carpooled, walking, public transportation, taxi cab or other methods and bicycle.

Sub Question 5: Q5) Are there any trends in transportation mode usage over the years?

```
travel_data %>%
  group_by(year, mode_of_trans) %>%
  summarise(total_count = sum(count)) %>%
  ggplot(aes(x = year, y = total_count, color = mode_of_trans)) +
  geom_line() +
  labs(title = "Trends in Transportation Mode Usage Over the Years",
       x = "Year",
       y = "Total Count")
```

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




Conclusion : We can see a significant increment of driving alone and carpooled methods but others more likely to the same.