Fuel Efficiency vs. Horsepower

library(tidyverse)

## -- Attaching packages ----------------------------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.0 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.5  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts -------------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

Introduction

We study the relationship between sale volume and month.

Data

We use BC Assessment housing transaction data.

bca <- read\_rds("data/bca\_sm.rds")

Analysis

We drop observations from 2018, since we have data only until September 2018.

library(lubridate)

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':  
##   
## date

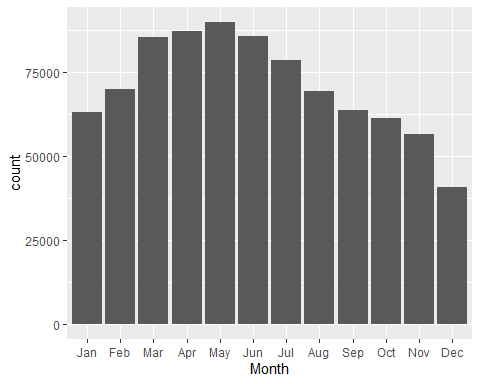
bca <- bca %>% filter(year(sale\_date) != 2018)

Our quick data analysis shows that sale volume is highest in May.

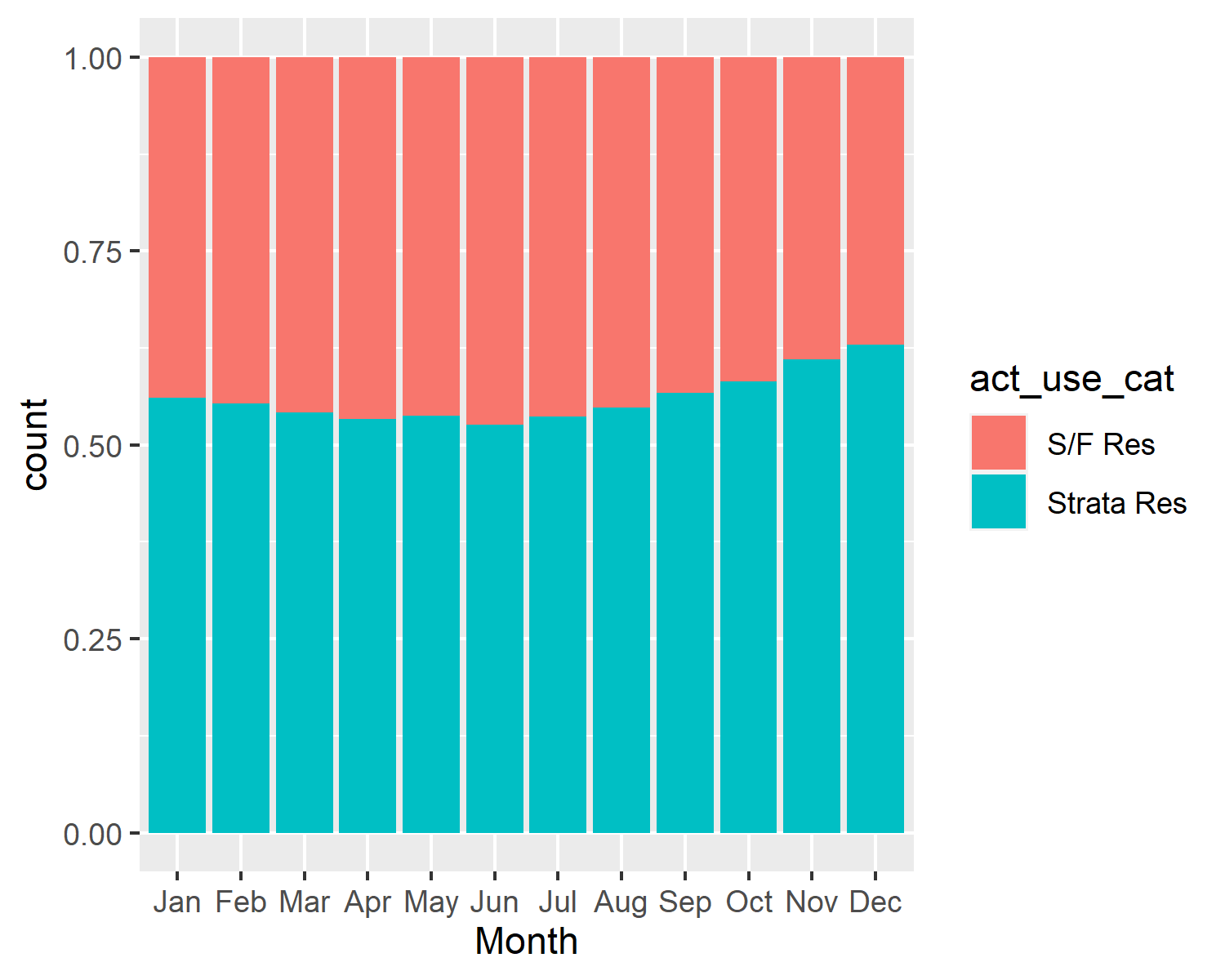
bca <- bca %>%   
 mutate(Month = month(sale\_date, label = TRUE))  
  
bca %>% count(Month)

## # A tibble: 12 x 2  
## Month n  
## <ord> <int>  
## 1 Jan 63215  
## 2 Feb 69903  
## 3 Mar 85413  
## 4 Apr 87428  
## 5 May 90049  
## 6 Jun 85856  
## 7 Jul 78696  
## 8 Aug 69368  
## 9 Sep 63645  
## 10 Oct 61426  
## 11 Nov 56761  
## 12 Dec 40894

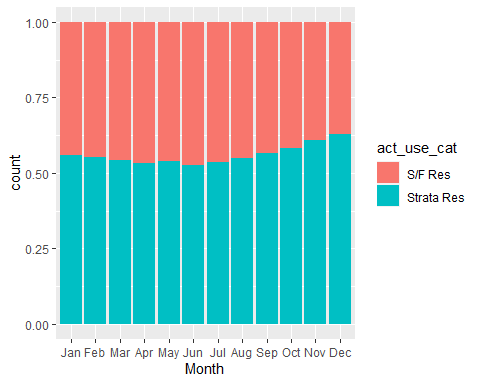
bca %>% ggplot() + geom\_bar(aes(Month))



bca %>% ggplot() + geom\_bar(aes(Month, fill=act\_use\_cat), position = "fill")



bca %>% ggplot() + geom\_bar(aes(Month, fill=act\_use\_cat), position = "fill")



Conclusion

The analysis shows that housing transaction volumes are higher from March to August. This is true for both single family detached houses and condos. The percentage of condos among housing transactions are higher from October to December.