GROUP PROJECT CSC 3220

2022-11-08

Load the Relevant Packages

```
state_market.df <- read.table("../data/state_market_tracker.tsv000", sep = '\t', header = TRUE)
```

```
#new_state_market.df = subset(state_market.df, select = -c(period_duration, region_type_id, table_id, i

#specific_date <- c("2017-12-31")
#specific_date <- as.Date(specific_date, format = "%Y-%m-%d")

#new_state_market.df = subset(new_state_market.df, period_begin > specific_date)
```

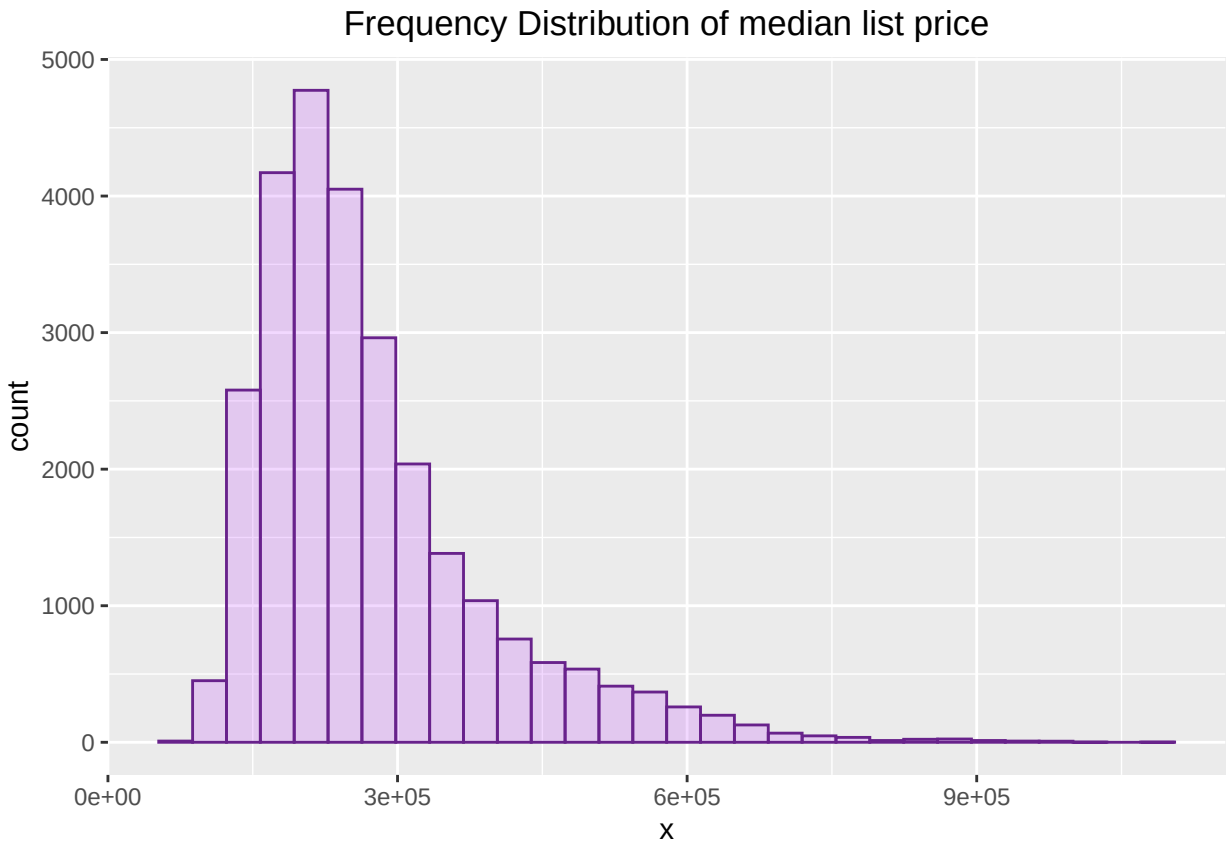
```
#dim(new_state_market.df)
#renderDataTable({datatable(new_state_market.df, options=list(scrollX=TRUE))})
#datatable(new_state_market.df, options=list(scrollX=TRUE))
```

```
mean_years <- data.frame(state_market.df$period_begin, state_market.df$median_sale_price_yoy, state_mark
```

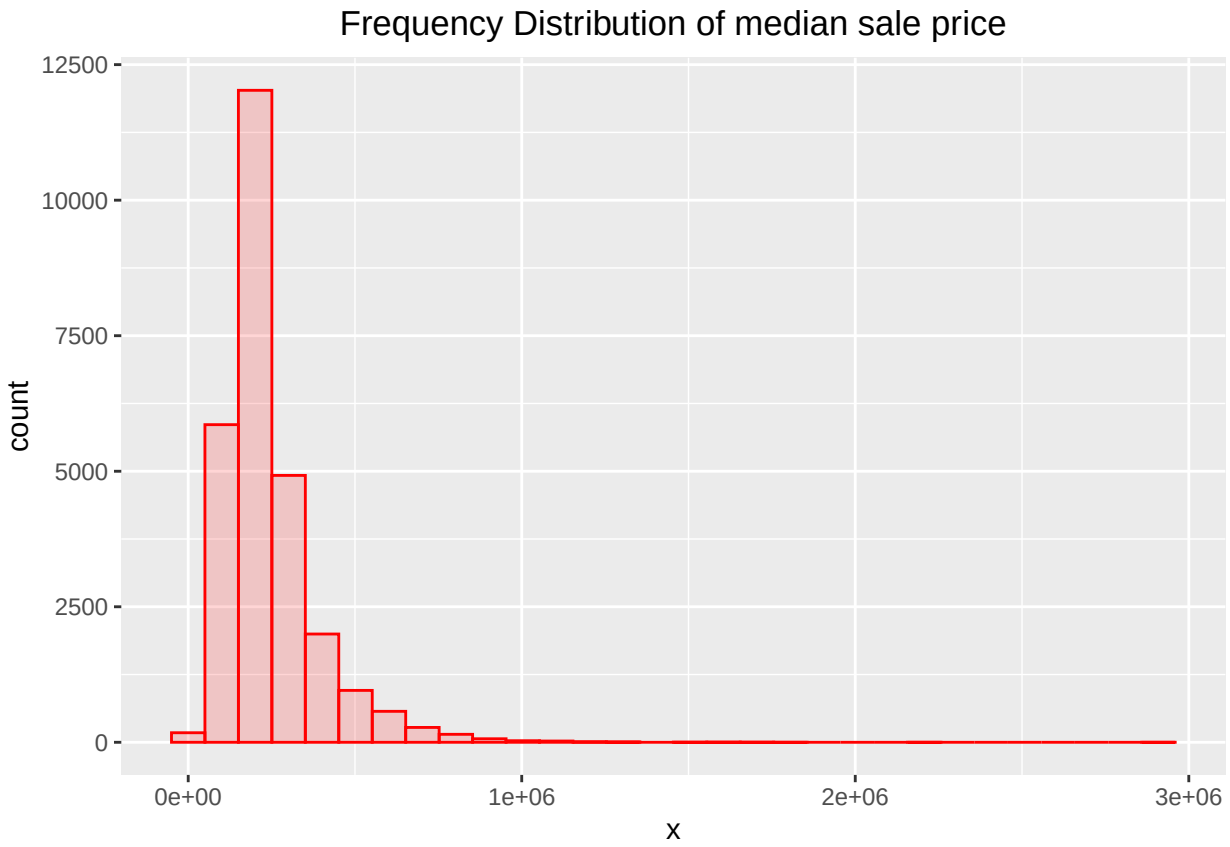
```
new_df <- subset.data.frame(state_market.df, select = c(state_code, median_list_price, median_sale_price

print(
    ggplot(new_df, aes_string(x=new_df$median_list_price))
    + geom_histogram(
      colour="darkorchid4", fill="darkorchid1", position="identity", bins=30, alpha=0.2
    )
    + ggtitle(paste("Frequency Distribution of median list price", sep=""))
    + theme(plot.title=element_text(hjust = 0.5)))
```

```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation ideoms with `aes()`
```

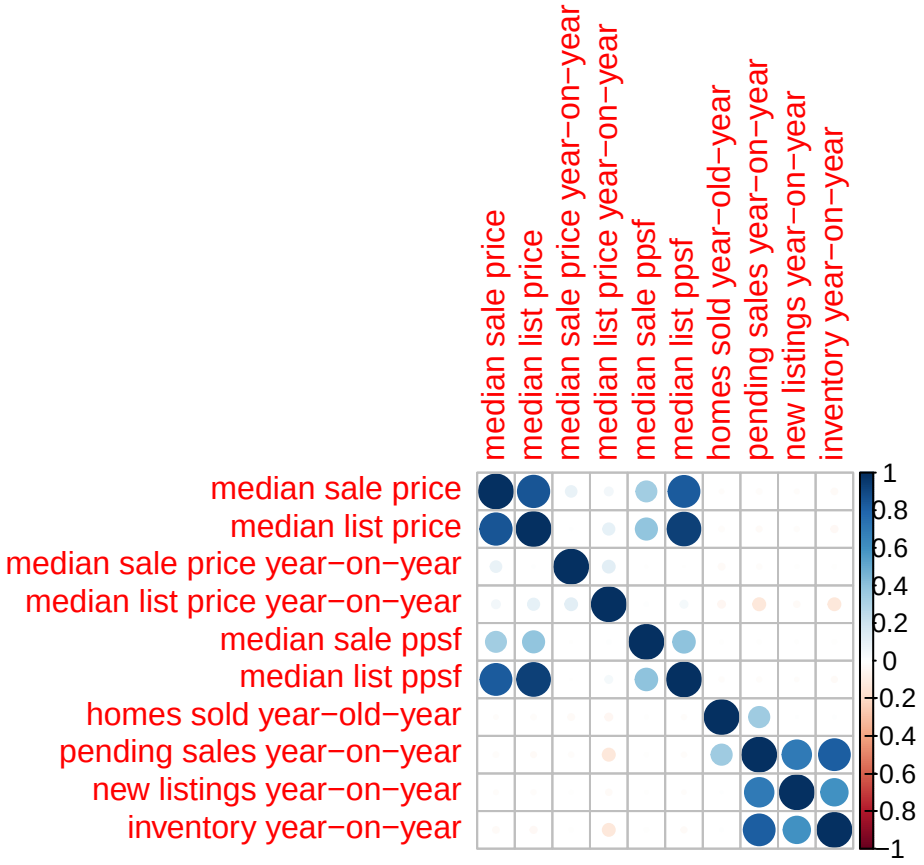
```
## Warning: Removed 142 rows containing non-finite values (`stat_bin()`).
```




```
print(
    ggplot(new_df, aes_string(x=new_df$median_sale_price))
    + geom_histogram(
      colour="red", fill="firebrick1", position="identity", bins=30, alpha=0.2
    )
    + ggtitle(paste("Frequency Distribution of median sale price", sep=""))
    + theme(plot.title=element_text(hjust = 0.5)))
```




```
cor.df <- subset.data.frame(state_market.df, select = c(median_sale_price, median_list_price, median_sa

cor.table <- cor(cor.df, use="pairwise.complete.obs")

rownames(cor.table) <- c("median sale price", "median list price", "median sale price year-on-year", "m

colnames(cor.table) <- c("median sale price", "median list price", "median sale price year-on-year", "m

matrix <- corrplot(cor.table)

corrplot(cor.table)
```

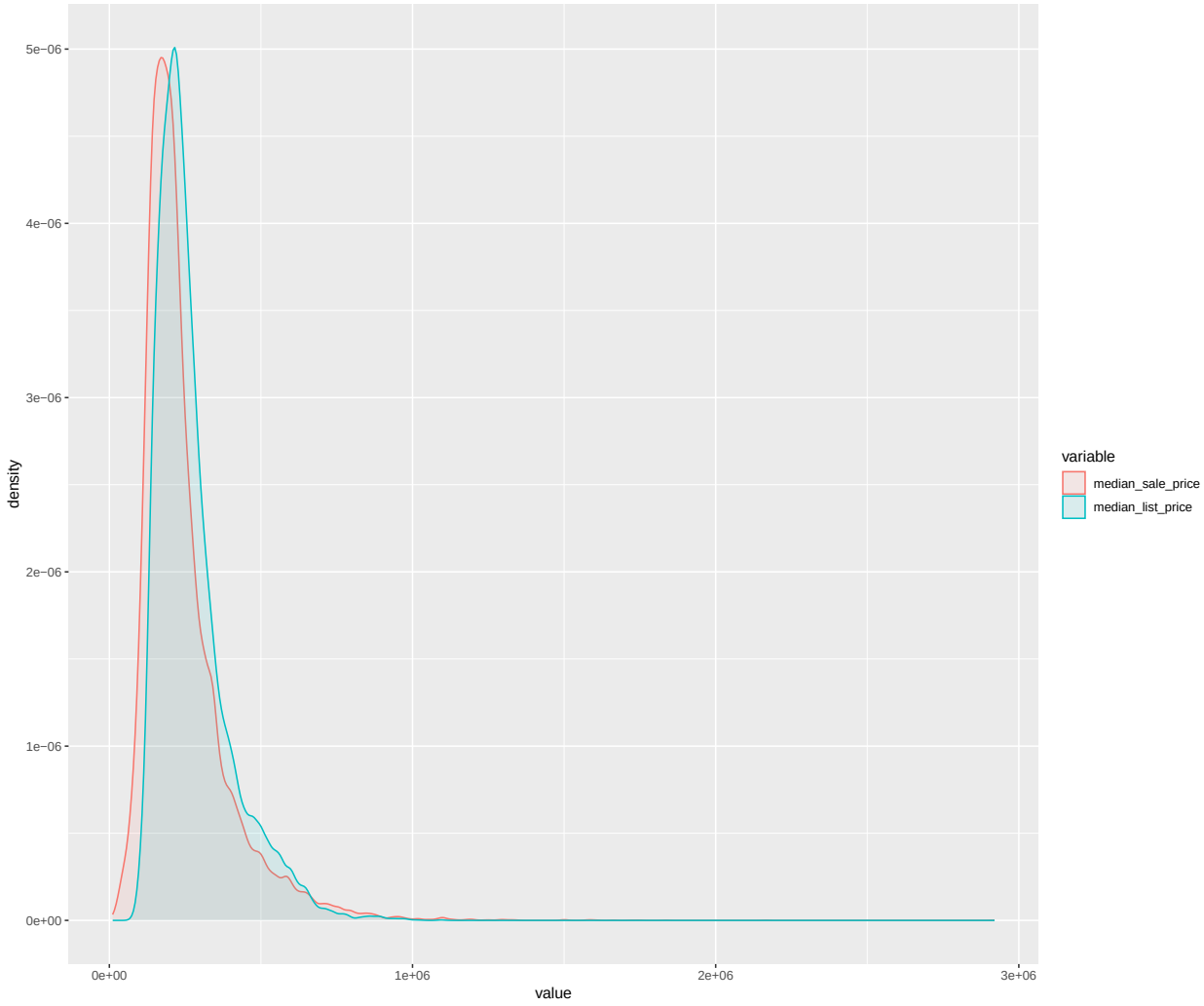



```
data_median_salemedian_list <- data.frame(state_market.df[, c("median_sale_price", "median_list_price")]

data_median_salemedian_list$refseq <- c("median_sale_price", "median_list_price")
s.plot <- melt(data_median_salemedian_list)
```

```
## Using refseq as id variables
```

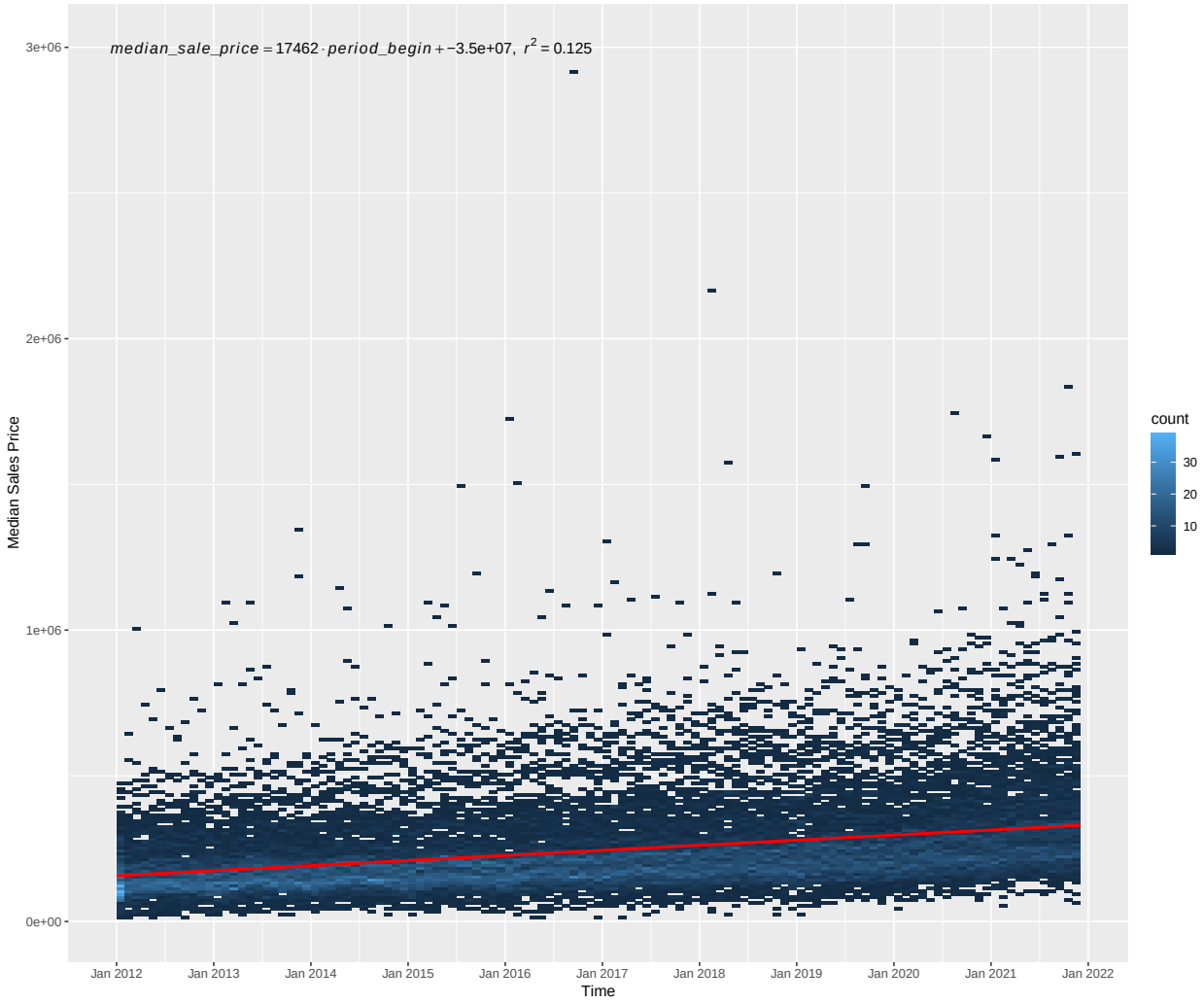
```
ggplot(s.plot, aes(x = value, colour = variable, fill = variable)) + geom_density(alpha = 0.1)
```

```
## Warning: Removed 142 rows containing non-finite values (‘stat_density()‘).
```




```
lm_eqn <- function(df){
    m <- lm(median_sale_price ~ as.yearmon(period_begin), df);
    eq <- substitute(italic(median_sale_price) == b %.% italic(period_begin) + a*","~~italic(r)^2~"="~r
         list(a = format(unname(coef(m)[1]), digits = 2),
              b = format(unname(coef(m)[2]), digits = 2),
             r2 = format(summary(m)$r.squared, digits = 3)))
    as.character(as.expression(eq));
}

ggplot(state_market.df, aes(x = as.yearmon(period_begin), y = median_sale_price)) +
                        geom_bin_2d(binwidth = c(1/12, 10000)) +
                        xlab("Time") +
                        ylab("Median Sales Price") +
                        scale_x_yearmon(n = 10) +
                        geom_smooth(method = "lm", se = FALSE, color = "red") +
                        annotate("text", x = as.numeric(as.yearmon("2014-06-01")), y = 3e+06, parse = T
```

```
## `geom_smooth()` using formula = 'y ~ x'
```


median_sale_price = 17462 · period_begin + −3.5e+07, r² = 0.125
3e+06
2e+06
1e+06
0e+00
Median Sales Price
Jan 2012
Jan 2013
Jan 2014
Jan 2015
Jan 2016
Jan 2017
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Time
count
30
20
10