GROUP PROJECT CSC 3220

2022-11-08

Load the Relevant Packages

```r
state_market.df <- read.table("../data/state_market_tracker.tsv000",
    sep = "\t", header = TRUE)
```

```r
# new_state_market.df = subset(state_market.df, select
# = -c(period_duration, region_type_id, table_id,
# is_seasonally_adjusted, region_type, city,
# parent_metro_region_metro_code, price_drops,
# price_drops_mom, price_drops_yoy))

# 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)
```

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

```r
state_market.df$period_begin <- as.Date(state_market.df$period_begin)
split_by_year <- split(state_market.df, format(state_market.df$period_begin,
    "%Y"))

lapply(split_by_year, function(i) ggplot(i, aes(x = state_code,
    y = median_sale_price, color = median_sale_price)) +
    geom_boxplot(colour = "firebrick1", fill = "orange",
        alpha = 5) + geom_point() + xlab(format(i$period_begin,
    "%Y")) + ylab("Median Sales Price") + scale_y_continuous(labels = function(x) format(x,
    scientific = TRUE)))
```
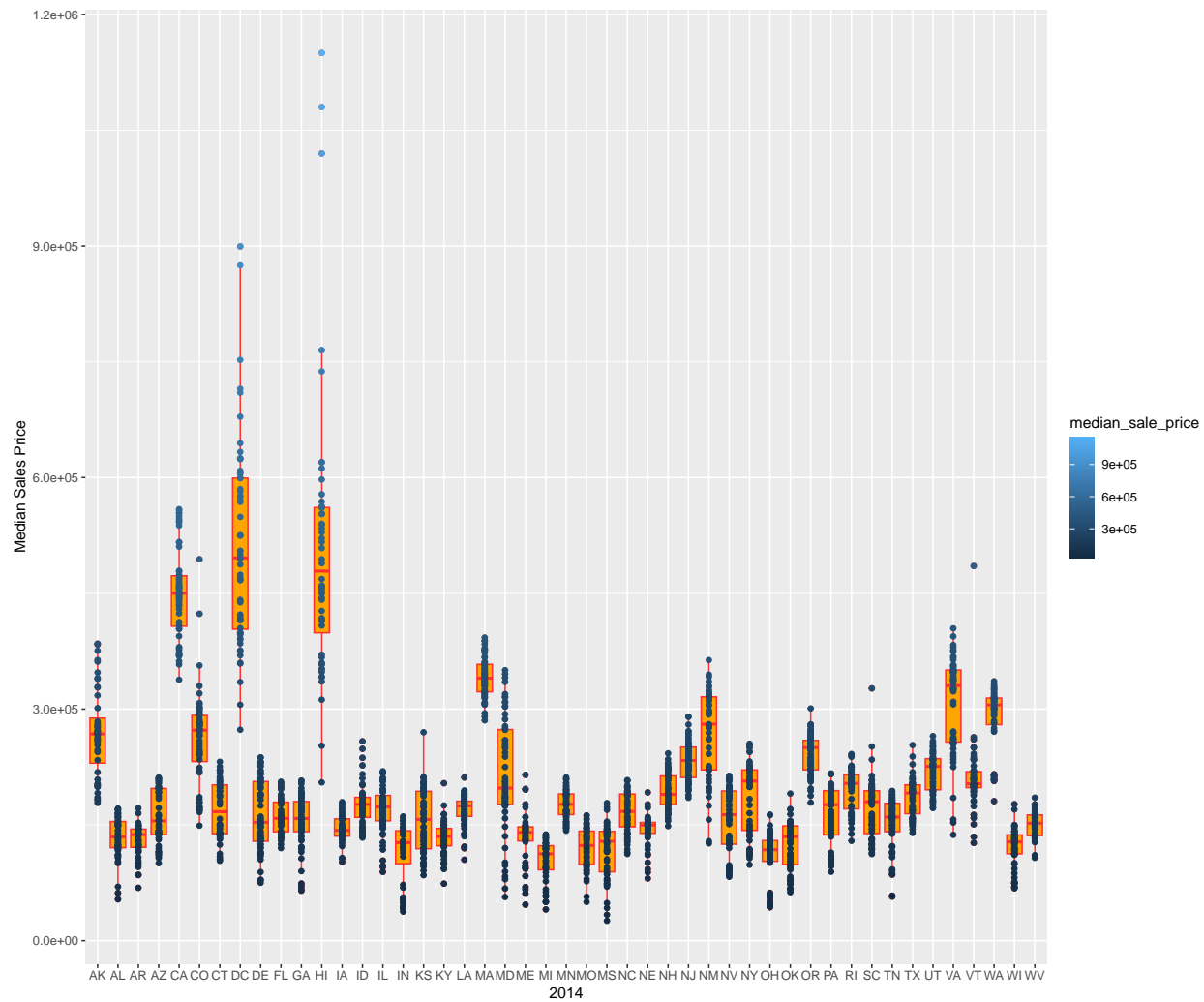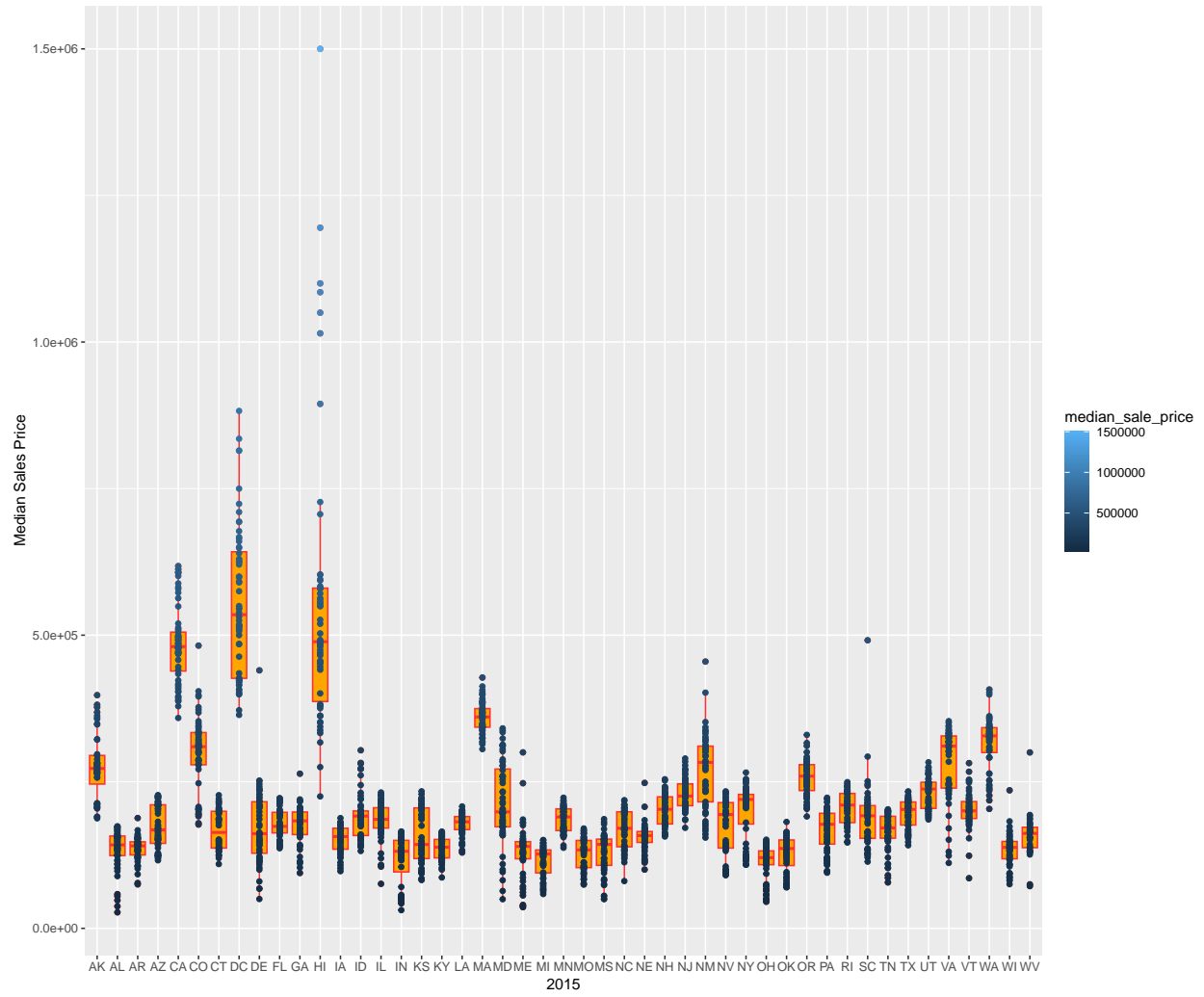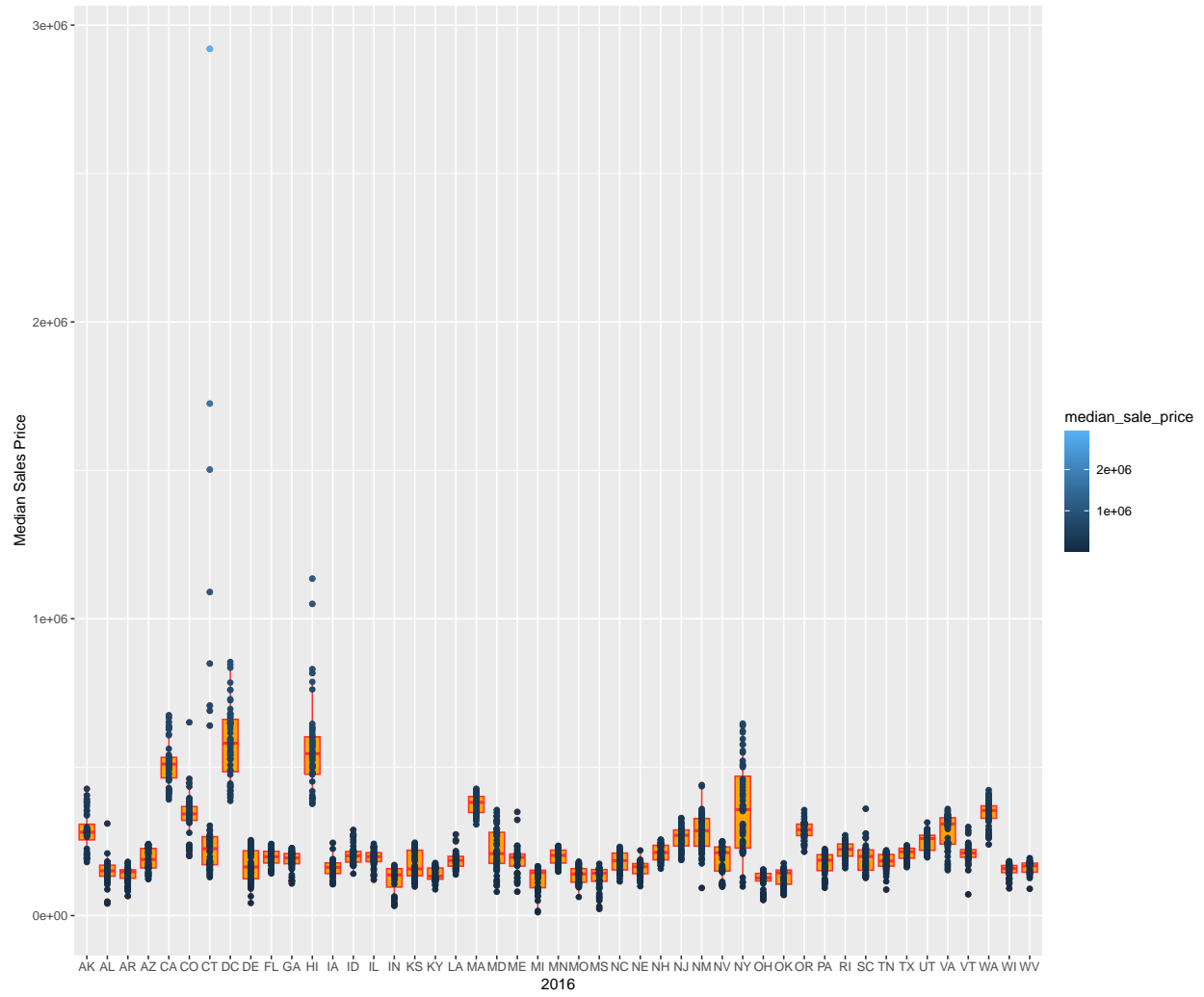
```
## $`2012`
```


$`2013`

Median Sales Price vs 2013 (boxplot by state)
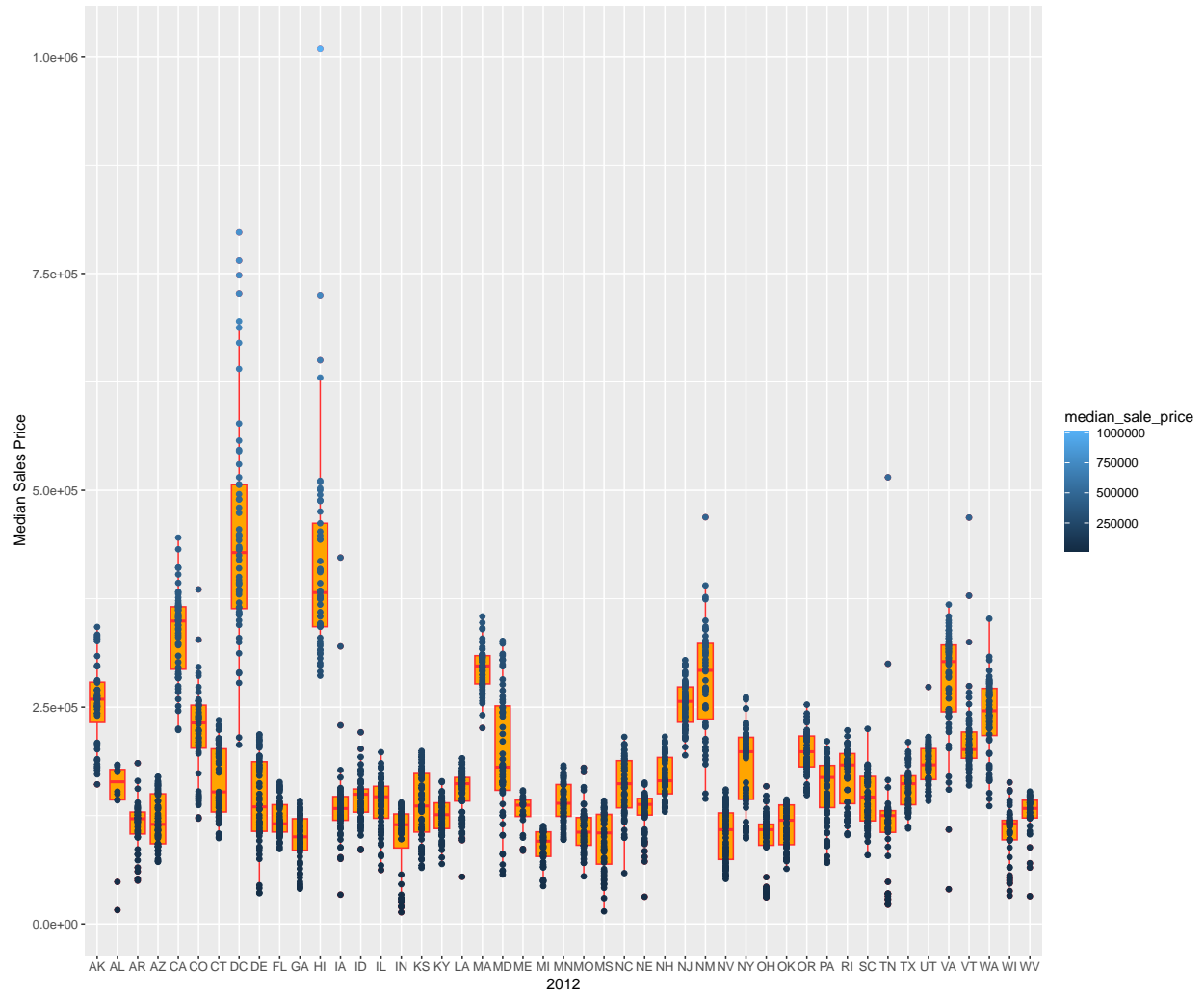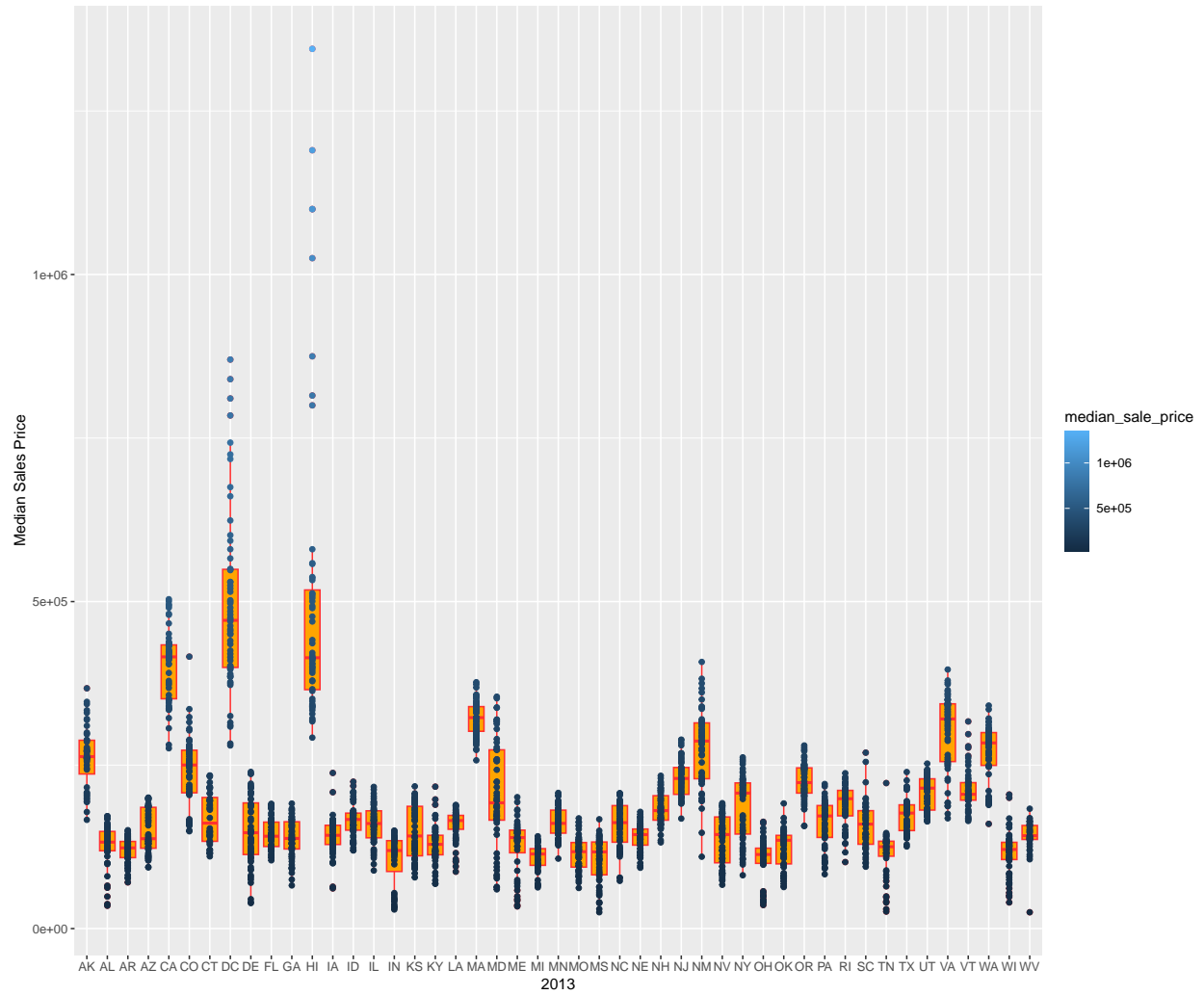
```
##
## $`2014`
```


$`2015`

$`2016`

$`2017`

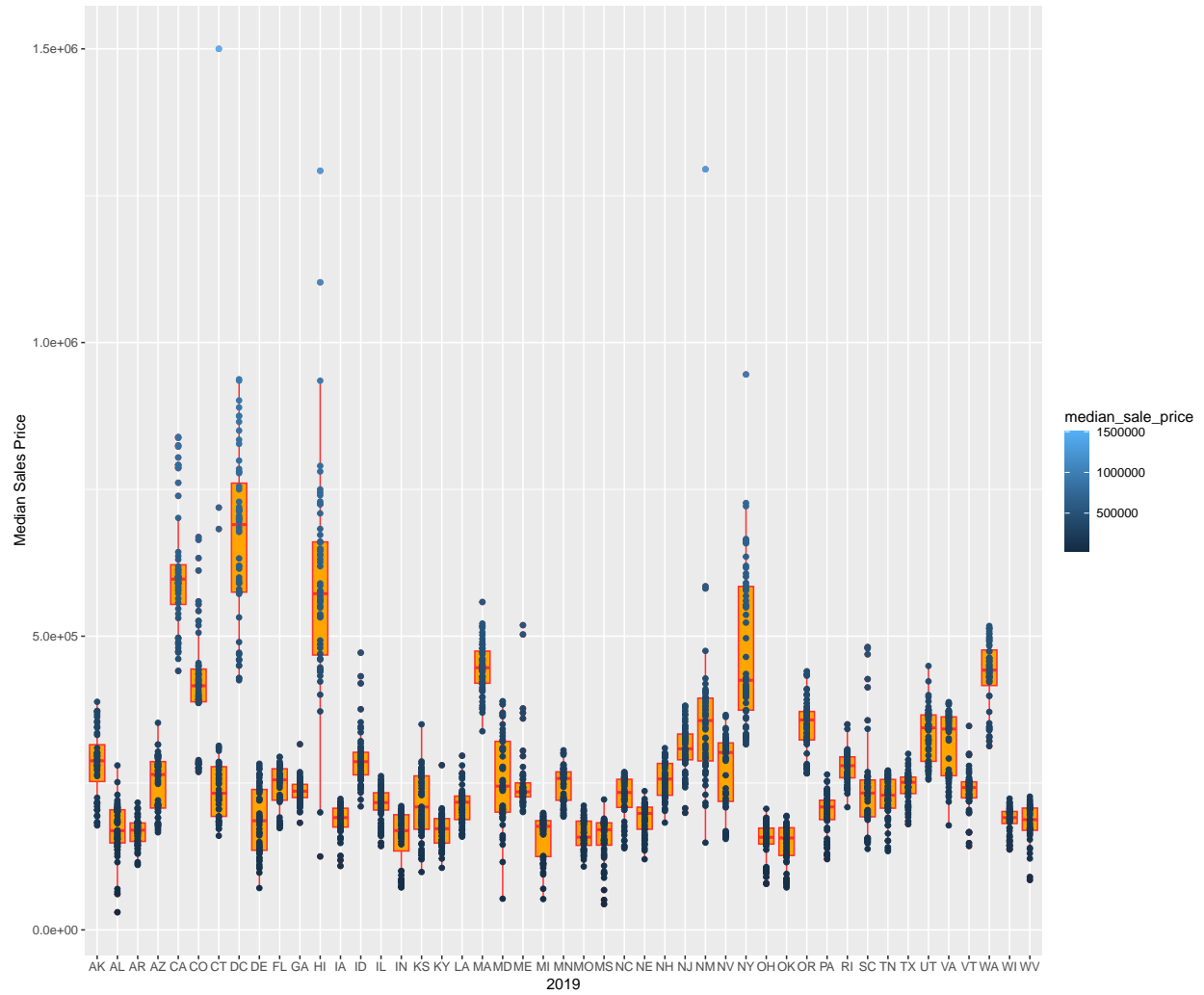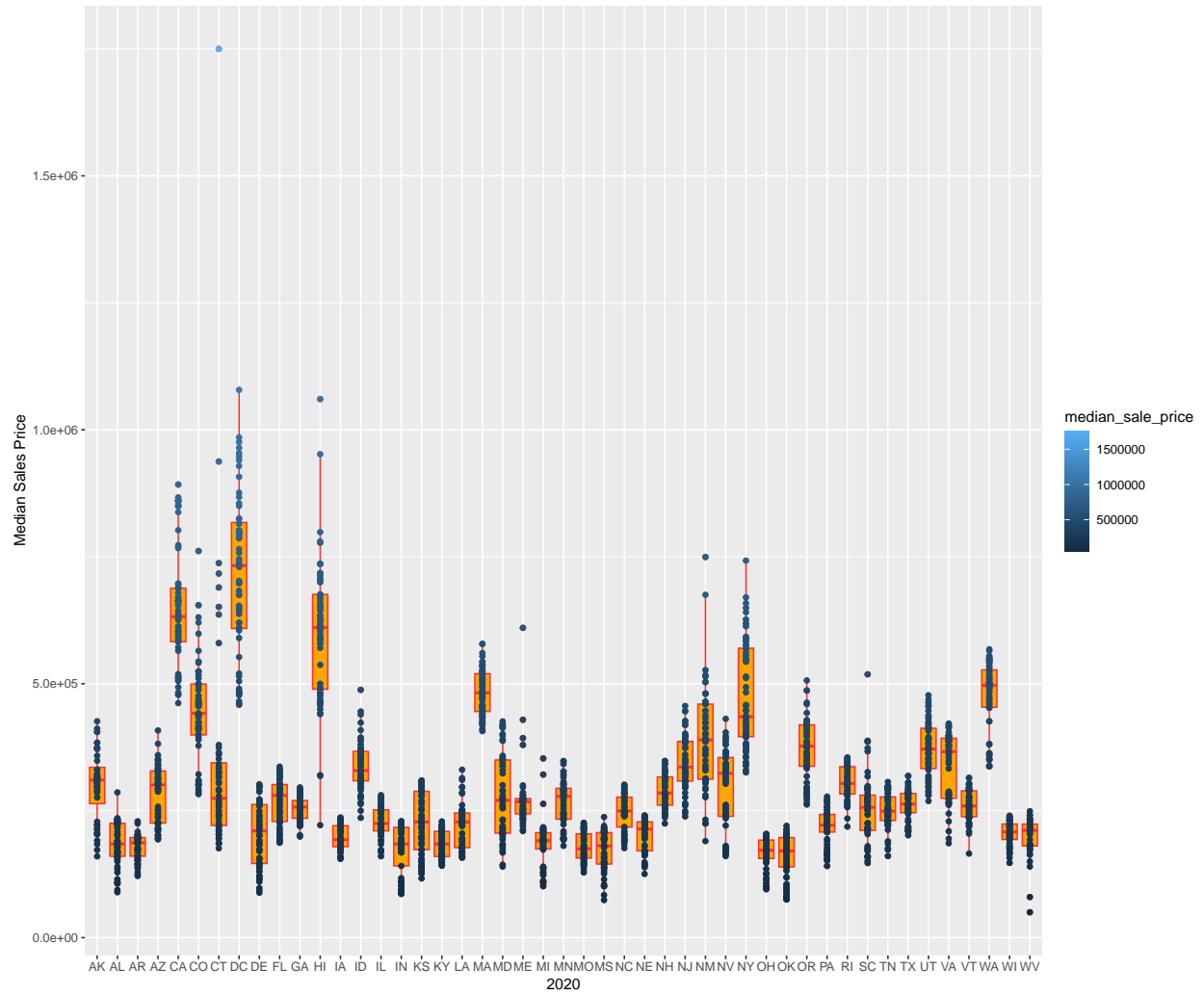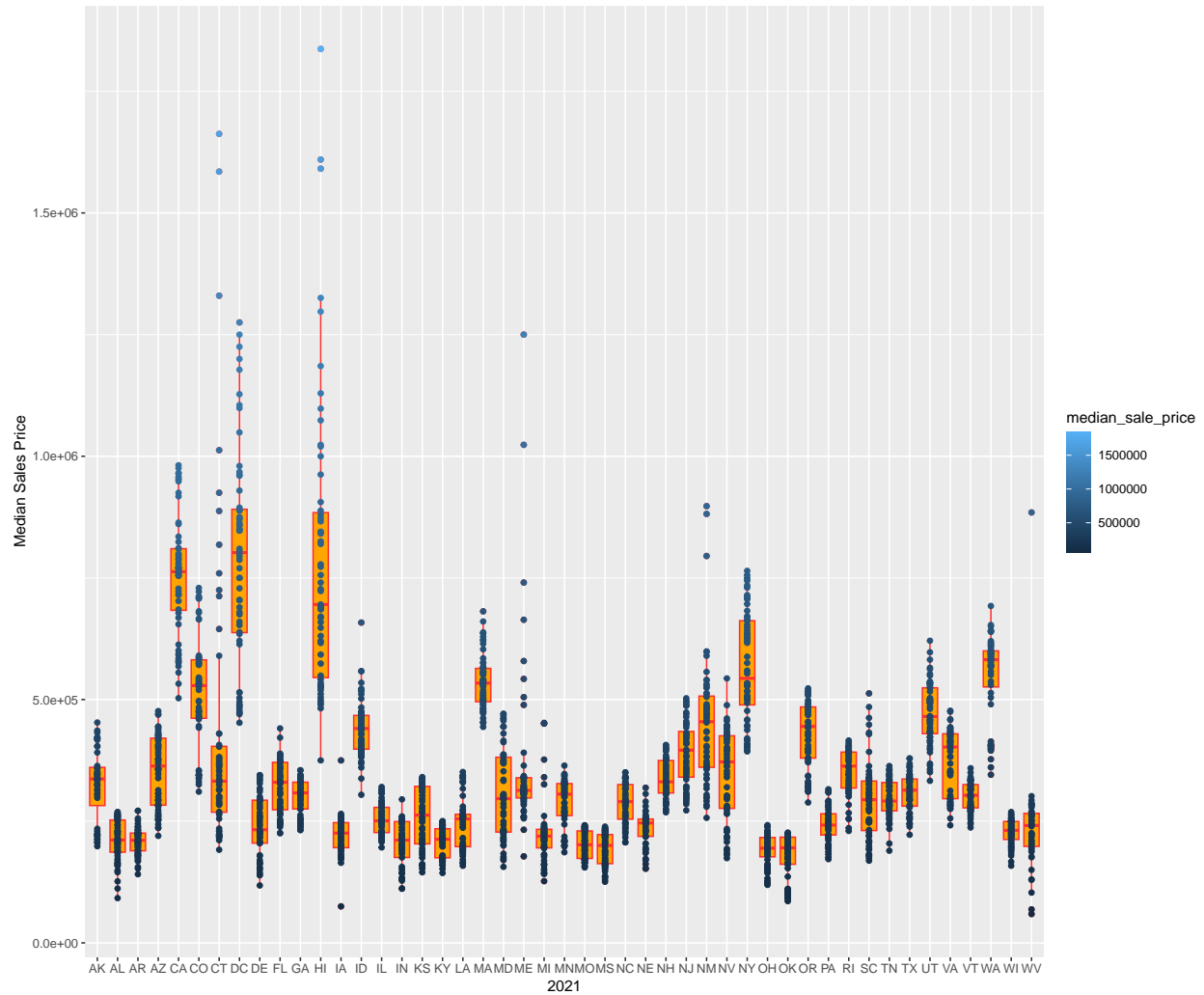## $`2018`

$`2019`

$`2020`

9
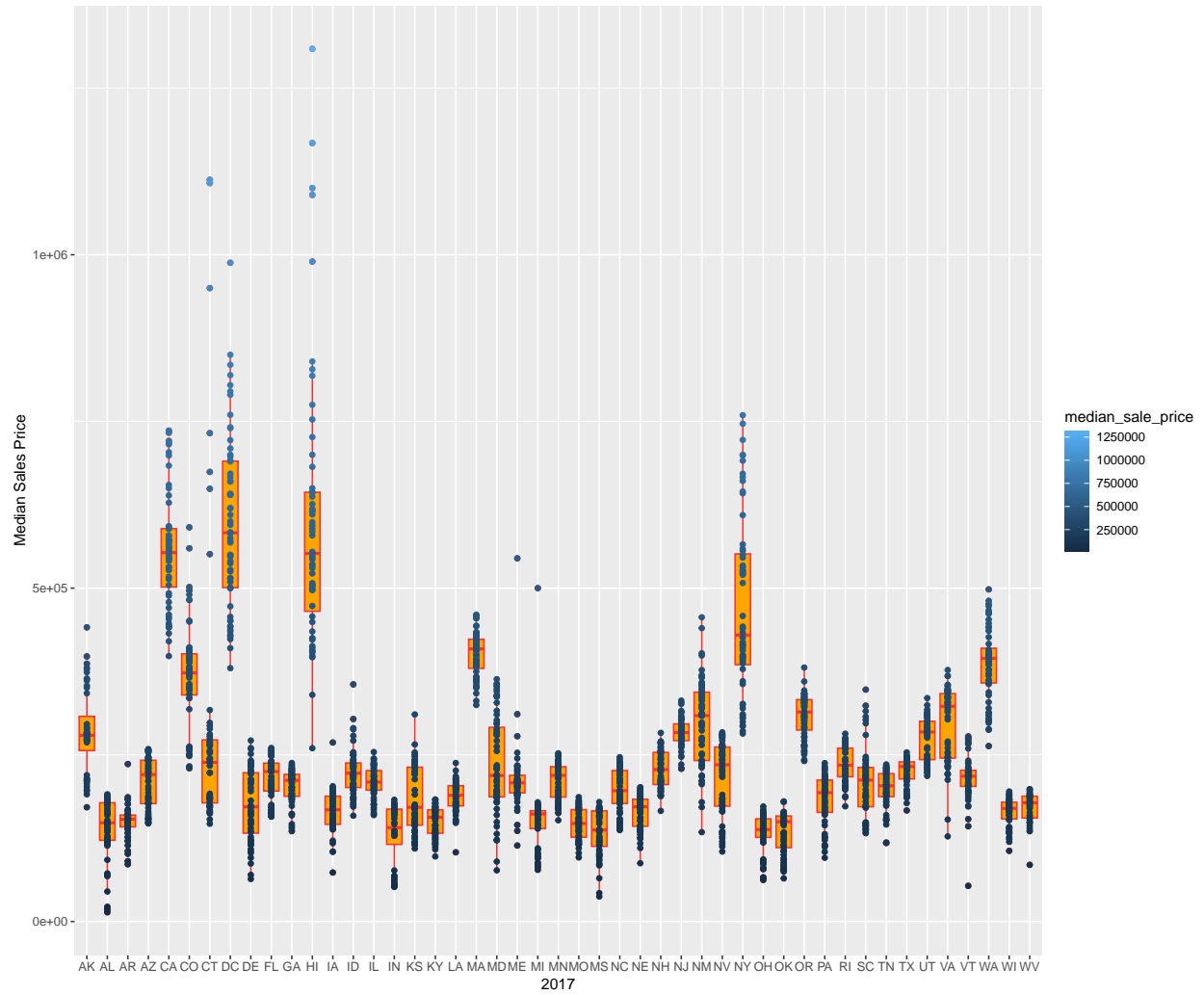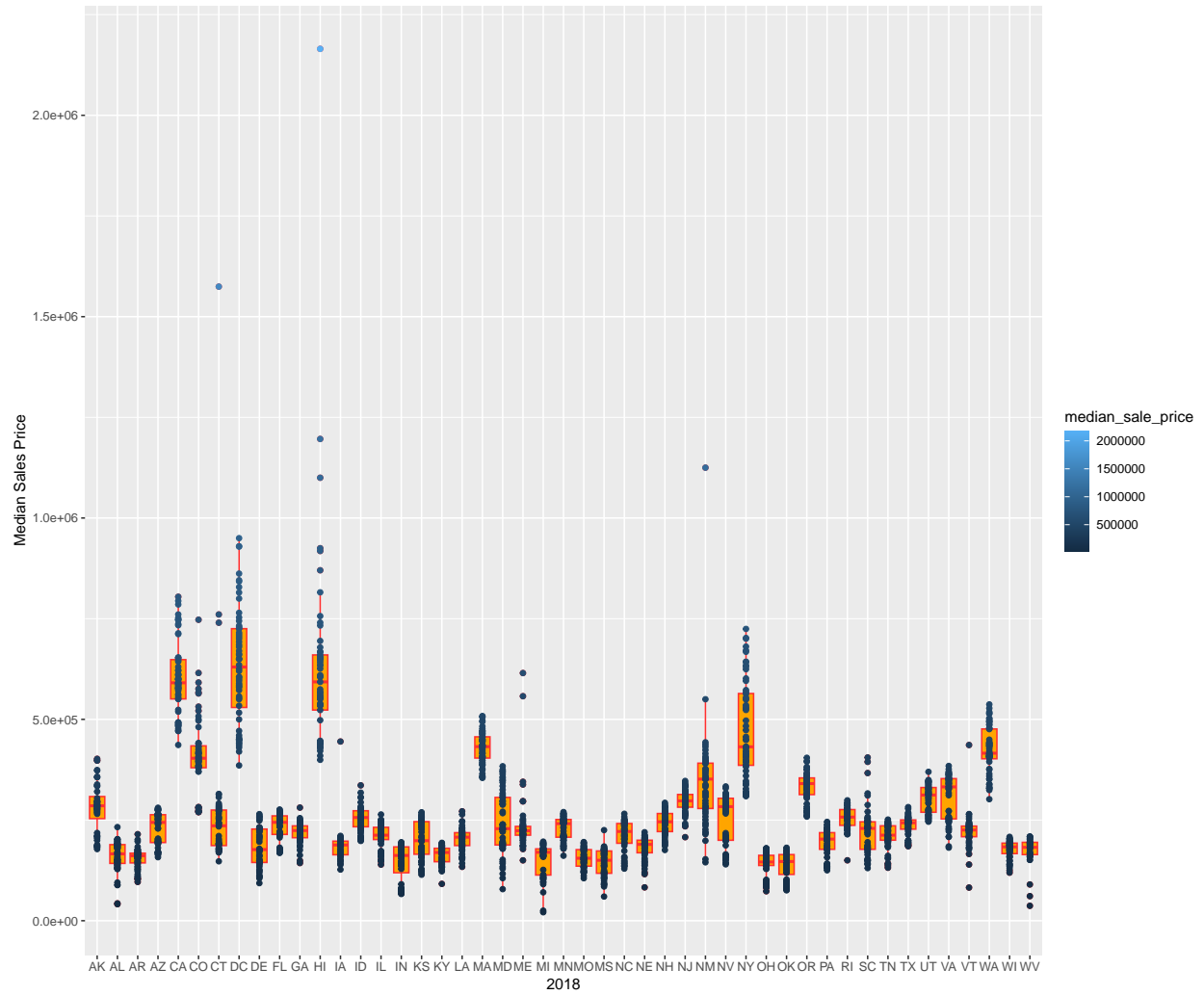
$`2021`

```
mean_years <- data.frame(state_market.df$period_begin, state_market.df$median_sale_price_yoy,
    state_market.df$state)
```
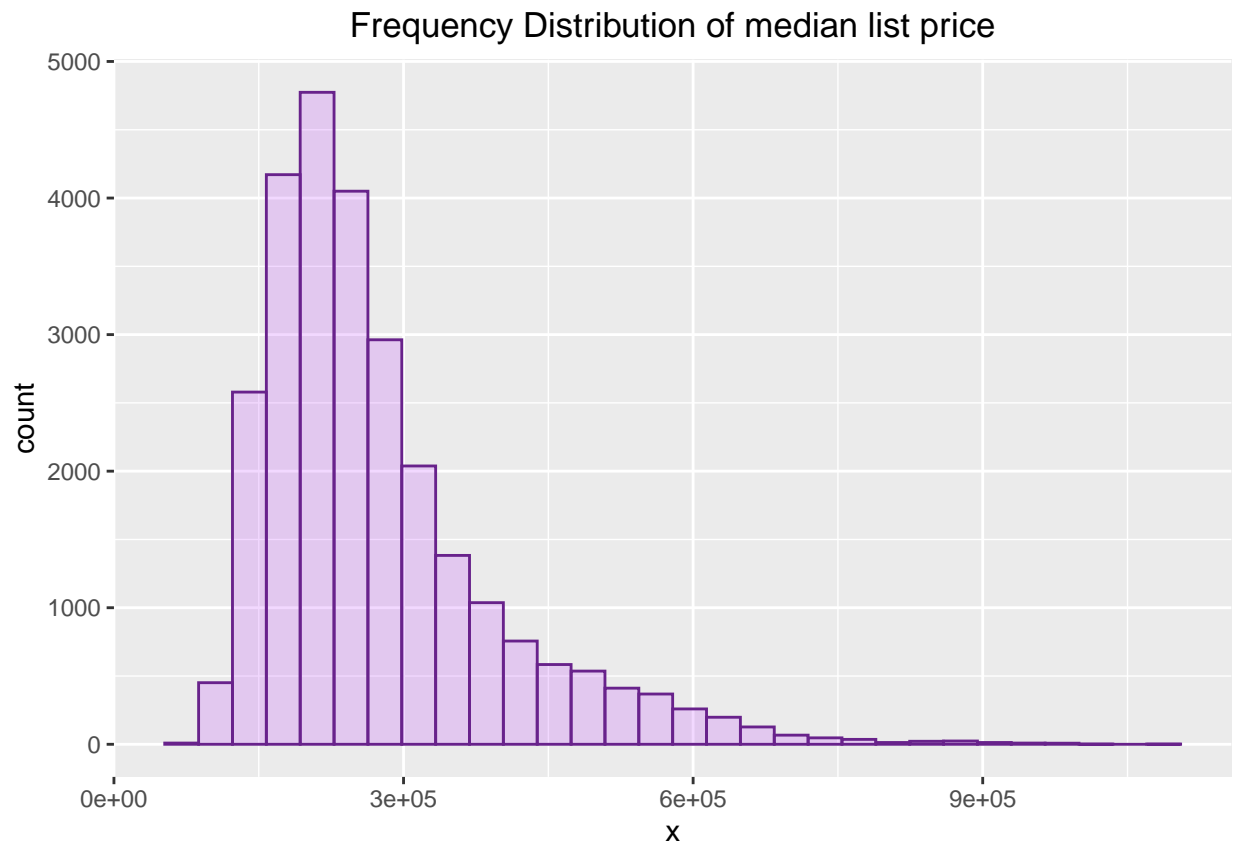
```
new_df <- subset.data.frame(state_market.df, select = c(state_code,
    median_list_price, median_sale_price), drop = FALSE)

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()').
```

Frequency Distribution of median list price


```
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)))
```
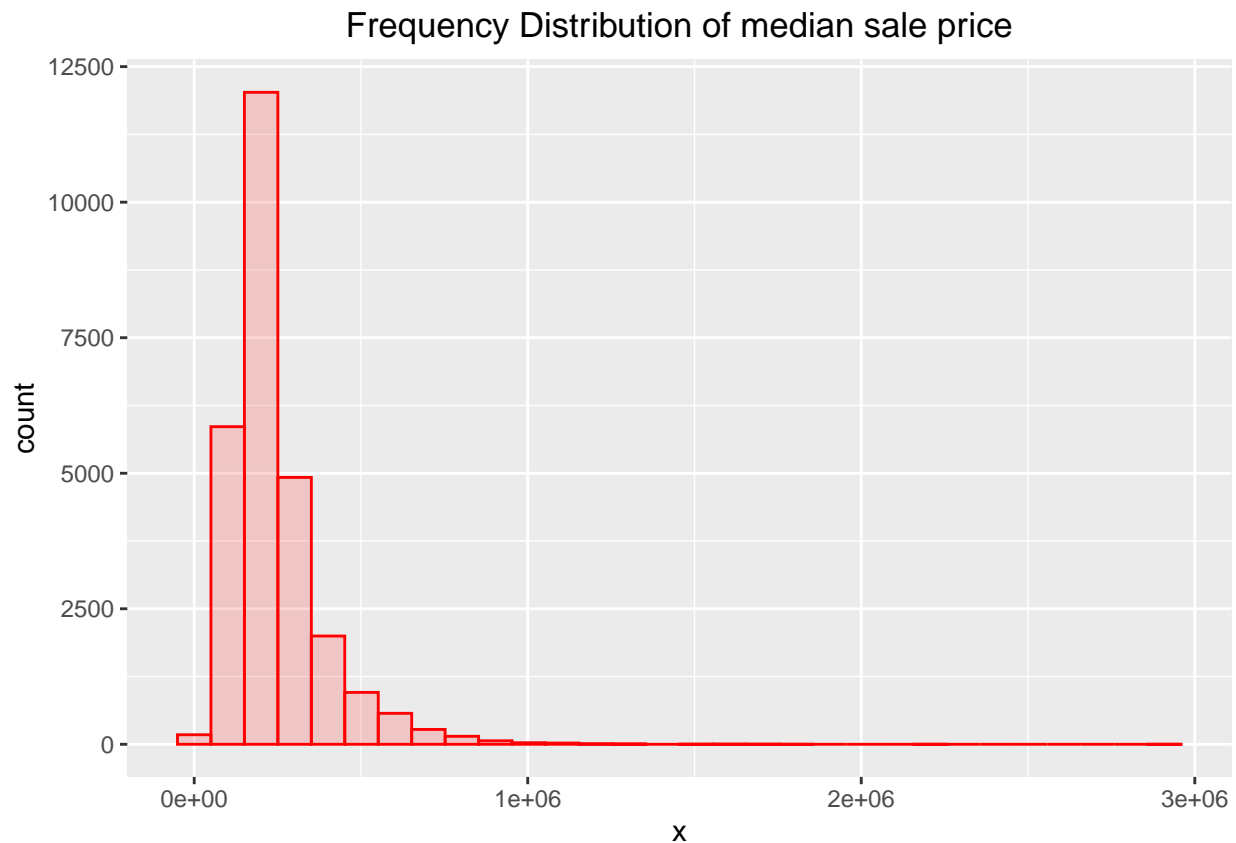
Frequency Distribution of median sale price


```
cor.df <- subset.data.frame(state_market.df, select = c(median_sale_price,
    median_list_price, median_sale_price_yoy, median_list_price_yoy,
    median_ppsf, median_list_ppsf, homes_sold_yoy, pending_sales_yoy,
    new_listings_yoy, inventory_yoy), drop = FALSE)


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", "median list price year-on-year",
    "median sale ppsf", "median list ppsf", "homes sold year-old-year",
    "pending sales year-on-year", "new listings year-on-year",
    "inventory year-on-year")

colnames(cor.table) <- c("median sale price", "median list price",
    "median sale price year-on-year", "median list price year-on-year",
    "median sale ppsf", "median list ppsf", "homes sold year-old-year",
    "pending sales year-on-year", "new listings year-on-year",
    "inventory year-on-year")

matrix <- corrplot(cor.table)

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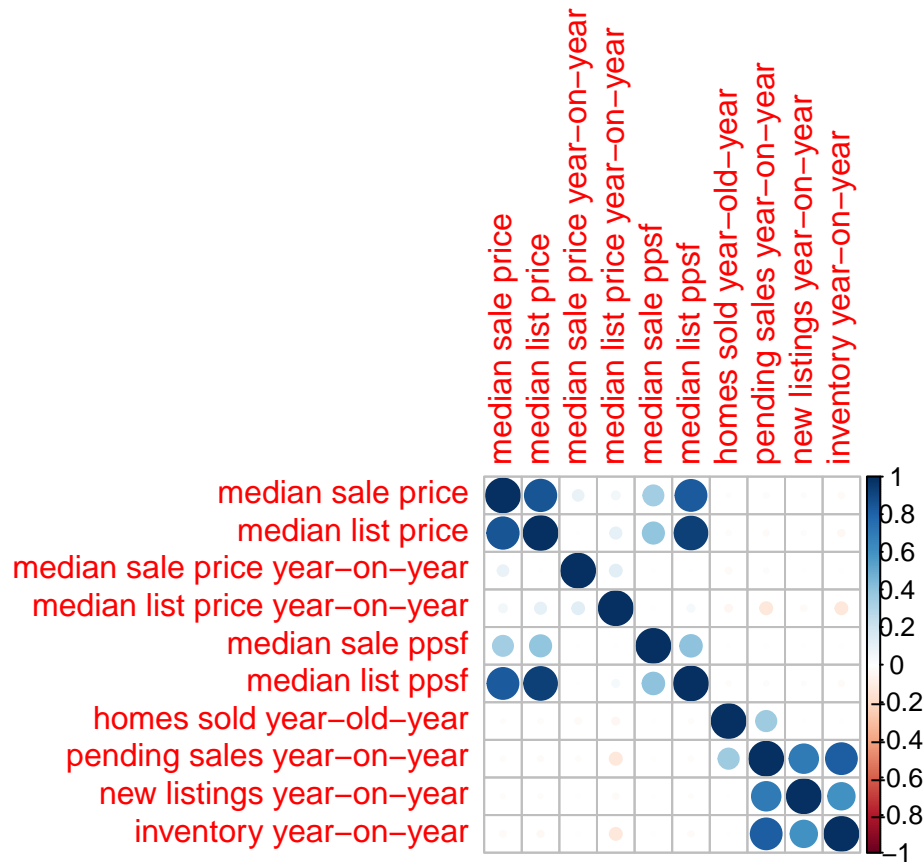
```r
data_median_salemedian_list <- data.frame(state_market.df[,
    c("median_sale_price", "median_list_price")])

data_median_salemedian_list$refseq <- c("median_sale_price",
    "median_list_price")
s.plot <- melt(data_median_salemedian_list)
```
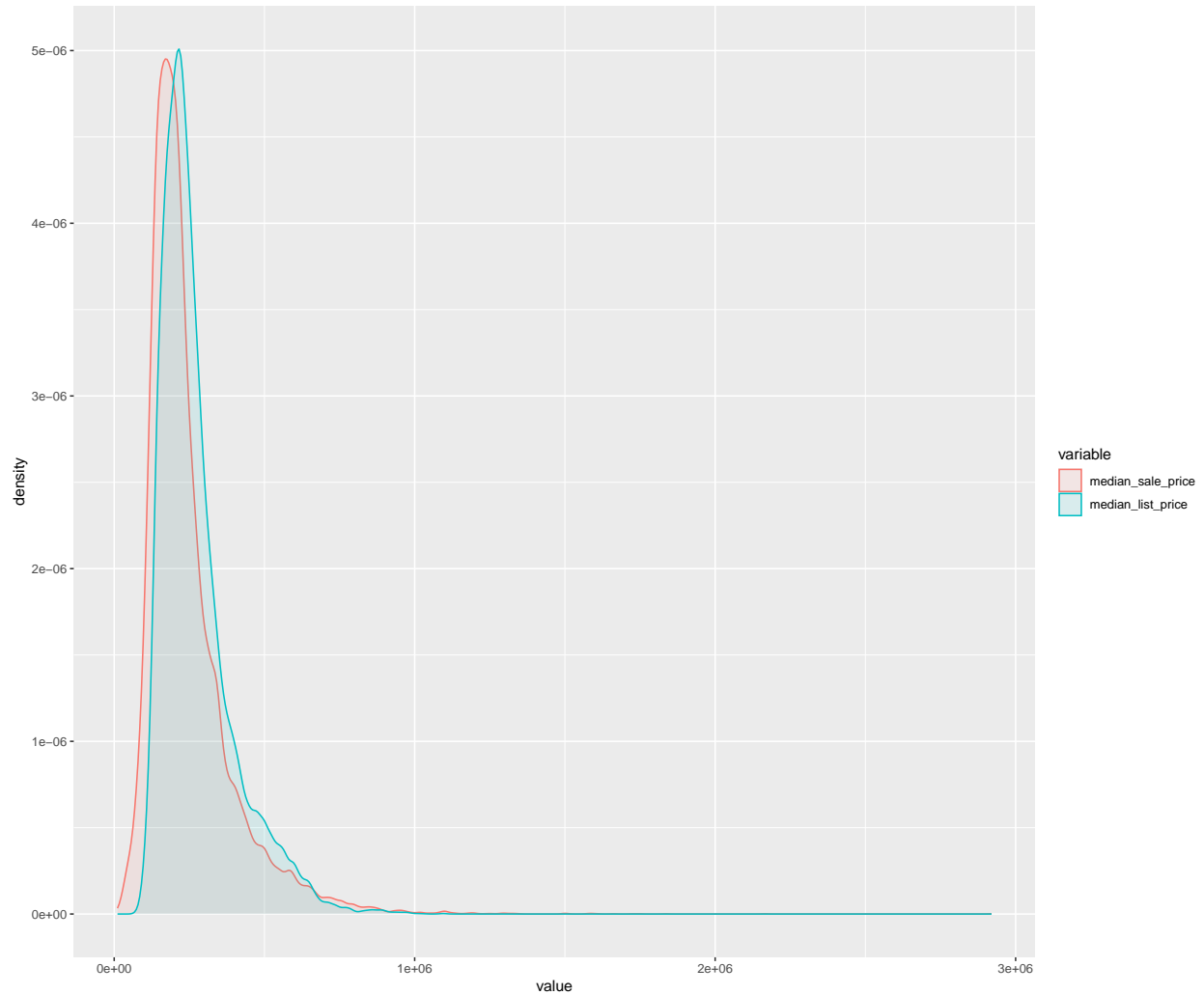
```
## Using refseq as id variables
```

```r
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()').
```

```r
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 ~ "=" ~ r2, 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 = TRUE, label = lm_eqn(state_market.df))
```

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

$median\_sale\_price = 17462 \cdot period\_begin + -3.5e+07, \ r^2 = 0.125$