

Economic Indicators analysis

2025-01-03

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Note to Reader

I recommend skipping to page 8 of this PDF as everything until then is just me exploring my dataset's columns and values.

Dataset analysis

```
all_econ_data = read.csv('./Key_Economic_Indicators.csv')
colnames(all_econ_data)
```

```
## [1] "Month"
## [2] "Year"
## [3] "Consumer.Confidence.Index.TX"
## [4] "Consumer.Confidence.West.South.Central"
## [5] "Consumer.Confidence.Index.US"
## [6] "PCE.Deflator"
## [7] "Consumer.Price.Index.TX"
## [8] "Consumer.Price.Index.U.S."
## [9] "CPI.U.S..Ex.Food.and.Energy"
## [10] "Nonfarm.Employment.TX"
## [11] "Nonfarm.Employment.U.S."
## [12] "Unemployment.TX"
## [13] "Unemployment.U.S."
## [14] "Single.Family.Building.Permits.TX"
## [15] "Multi.Family.Building.Permits.TX"
## [16] "Existing.Single.Family.Home.Sales.TX"
## [17] "Existing.Single.Family.Home.Price.TX"
## [18] "Non.Residential.Building.Construction"
## [19] "Total.Sales.Tax.Collections.Retail.TX"
## [20] "Total.Sales.Tax.Collections.TX"
## [21] "Retail.Gasoline.Price.TX"
## [22] "Retail.Diesel.Price.TX"
```

```
## [23] "Nonfarm.Employment.Illinois"
## [24] "Nonfarm.Employment.Florida"
## [25] "Nonfarm.Employment.New.York"
## [26] "Nonfarm.Employment.Texas"
## [27] "Nonfarm.Employment.California"
## [28] "Gross.Value.Crude.Oil.Production"
## [29] "Gross.Value.Natural.Gas.Production"
## [30] "Motor.Fuel.Taxed.Gasoline"
## [31] "Motor.Fuel.Taxed.Diesel"
```

```
head(all_econ_data)
```

```
##      Month Year Consumer.Confidence.Index.TX
## 1      1 2005                               NA
## 2      2 2005                               NA
## 3      3 2005                               NA
## 4      4 2005                               NA
## 5      5 2005                               NA
## 6      6 2005                               NA
##      Consumer.Confidence.West.South.Central Consumer.Confidence.Index.US
## 1                               NA                               NA
## 2                               NA                               NA
## 3                               NA                               NA
## 4                               NA                               NA
## 5                               NA                               NA
## 6                               NA                               NA
##      PCE.Deflator Consumer.Price.Index.TX Consumer.Price.Index.U.S.
## 1             NA                      NA                      NA
## 2             NA                      NA                      NA
## 3             NA                      NA                      NA
## 4             NA                      NA                      NA
## 5             NA                      NA                      NA
## 6             NA                      NA                      NA
##      CPI.U.S..Ex.Food.and.Energy Nonfarm.Employment.TX Nonfarm.Employment.U.S.
## 1                               NA                      NA                      NA
## 2                               NA                      NA                      NA
## 3                               NA                      NA                      NA
## 4                               NA                      NA                      NA
## 5                               NA                      NA                      NA
## 6                               NA                      NA                      NA
##      Unemployment.TX Unemployment.U.S. Single.Family.Building.Permits.TX
## 1             NA                      NA                      NA
## 2             NA                      NA                      NA
## 3             NA                      NA                      NA
## 4             NA                      NA                      NA
## 5             NA                      NA                      NA
## 6             NA                      NA                      NA
##      Multi.Family.Building.Permits.TX Existing.Single.Family.Home.Sales.TX
## 1                               NA                      NA
## 2                               NA                      NA
## 3                               NA                      NA
## 4                               NA                      NA
## 5                               NA                      NA
## 6                               NA                      NA
```

```

## Existing.Single.Family.Home.Price.TX Non.Residential.Building.Construction
## 1 NA NA
## 2 NA NA
## 3 NA NA
## 4 NA NA
## 5 NA NA
## 6 NA NA
## Total.Sales.Tax.Collections.Retail.TX Total.Sales.Tax.Collections.TX
## 1 NA NA
## 2 NA NA
## 3 NA NA
## 4 NA NA
## 5 NA NA
## 6 NA NA
## Retail.Gasoline.Price.TX Retail.Diesel.Price.TX Nonfarm.Employment.Illinois
## 1 1.773 1.906 5829.6
## 2 1.841 1.958 5836.2
## 3 2.008 2.148 5833.8
## 4 2.169 2.226 5857.5
## 5 2.088 2.150 5855.4
## 6 2.101 2.256 5841.0
## Nonfarm.Employment.Florida Nonfarm.Employment.New.York
## 1 7650.2 8490.6
## 2 7670.7 8487.9
## 3 7668.7 8483.6
## 4 7716.4 8518.6
## 5 7753.0 8509.9
## 6 7761.6 8515.9
## Nonfarm.Employment.Texas Nonfarm.Employment.California
## 1 9642.4 14881.8
## 2 9653.2 14908.3
## 3 9670.8 14929.3
## 4 9715.1 14979.4
## 5 9727.1 14985.4
## 6 9734.0 15000.9
## Gross.Value.Crude.Oil.Production Gross.Value.Natural.Gas.Production
## 1 NA NA
## 2 NA NA
## 3 NA NA
## 4 NA NA
## 5 NA NA
## 6 NA NA
## Motor.Fuel.Taxed.Gasoline Motor.Fuel.Taxed.Diesel
## 1 NA NA
## 2 NA NA
## 3 NA NA
## 4 NA NA
## 5 NA NA
## 6 NA NA

```

```

# Load necessary library
library(dplyr)

```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
# Arrange data by Year and Month for proper differencing
```

```
all_econ_data <- all_econ_data %>%
```

```
  arrange(Year, Month)
```

```
# Calculate month-over-month differences for all numeric columns
```

```
# Exclude 'Year' and 'Month' from differencing
```

```
data_diff <- all_econ_data %>%
```

```
  mutate(across(-c(Year, Month), ~ . - lag(.), .names = "diff_{col}"))
```

```
# View the resulting dataset
```

```
head(data_diff)
```

```
## Month Year Consumer.Confidence.Index.TX
```

```
## 1 1 2005 NA
```

```
## 2 2 2005 NA
```

```
## 3 3 2005 NA
```

```
## 4 4 2005 NA
```

```
## 5 5 2005 NA
```

```
## 6 6 2005 NA
```

```
## Consumer.Confidence.West.South.Central Consumer.Confidence.Index.US
```

```
## 1 NA NA
```

```
## 2 NA NA
```

```
## 3 NA NA
```

```
## 4 NA NA
```

```
## 5 NA NA
```

```
## 6 NA NA
```

```
## PCE.Deflator Consumer.Price.Index.TX Consumer.Price.Index.U.S.
```

```
## 1 NA NA NA
```

```
## 2 NA NA NA
```

```
## 3 NA NA NA
```

```
## 4 NA NA NA
```

```
## 5 NA NA NA
```

```
## 6 NA NA NA
```

```
## CPI.U.S..Ex.Food.and.Energy Nonfarm.Employment.TX Nonfarm.Employment.U.S.
```

```
## 1 NA NA NA
```

```
## 2 NA NA NA
```

```
## 3 NA NA NA
```

```
## 4 NA NA NA
```

```
## 5 NA NA NA
```

```
## 6 NA NA NA
```

```
## Unemployment.TX Unemployment.U.S. Single.Family.Building.Permits.TX
```

```
## 1 NA NA NA
```

```
## 2 NA NA NA
```

## 3	NA	NA	NA
## 4	NA	NA	NA
## 5	NA	NA	NA
## 6	NA	NA	NA
##	Multi.Family.Building.Permits.TX Existing.Single.Family.Home.Sales.TX		
## 1		NA	NA
## 2		NA	NA
## 3		NA	NA
## 4		NA	NA
## 5		NA	NA
## 6		NA	NA
##	Existing.Single.Family.Home.Price.TX Non.Residential.Building.Construction		
## 1		NA	NA
## 2		NA	NA
## 3		NA	NA
## 4		NA	NA
## 5		NA	NA
## 6		NA	NA
##	Total.Sales.Tax.Collections.Retail.TX Total.Sales.Tax.Collections.TX		
## 1		NA	NA
## 2		NA	NA
## 3		NA	NA
## 4		NA	NA
## 5		NA	NA
## 6		NA	NA
##	Retail.Gasoline.Price.TX Retail.Diesel.Price.TX Nonfarm.Employment.Illinois		
## 1	1.773	1.906	5829.6
## 2	1.841	1.958	5836.2
## 3	2.008	2.148	5833.8
## 4	2.169	2.226	5857.5
## 5	2.088	2.150	5855.4
## 6	2.101	2.256	5841.0
##	Nonfarm.Employment.Florida Nonfarm.Employment.New.York		
## 1	7650.2	8490.6	
## 2	7670.7	8487.9	
## 3	7668.7	8483.6	
## 4	7716.4	8518.6	
## 5	7753.0	8509.9	
## 6	7761.6	8515.9	
##	Nonfarm.Employment.Texas Nonfarm.Employment.California		
## 1	9642.4	14881.8	
## 2	9653.2	14908.3	
## 3	9670.8	14929.3	
## 4	9715.1	14979.4	
## 5	9727.1	14985.4	
## 6	9734.0	15000.9	
##	Gross.Value.Crude.Oil.Production Gross.Value.Natural.Gas.Production		
## 1		NA	NA
## 2		NA	NA
## 3		NA	NA
## 4		NA	NA
## 5		NA	NA
## 6		NA	NA
##	Motor.Fuel.Taxed.Gasoline Motor.Fuel.Taxed.Diesel		

## 1	NA	NA
## 2	NA	NA
## 3	NA	NA
## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
##	diff_Consumer.Confidence.Index.TX	diff_Consumer.Confidence.West.South.Central
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA
## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
##	diff_Consumer.Confidence.Index.US	diff_PCE.Deflator
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA
## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
##	diff_Consumer.Price.Index.TX	diff_Consumer.Price.Index.U.S.
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA
## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
##	diff_CPI.U.S..Ex.Food.and.Energy	diff_Nonfarm.Employment.TX
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA
## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
##	diff_Nonfarm.Employment.U.S.	diff_Unemployment.TX
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA
## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
##	diff_Single.Family.Building.Permits.TX	diff_Multi.Family.Building.Permits.TX
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA
## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
##	diff_Existing.Single.Family.Home.Sales.TX	
## 1	NA	
## 2	NA	
## 3	NA	
## 4	NA	
## 5	NA	

## 6	NA	
## diff_Existing.Single.Family.Home.Price.TX		
## 1	NA	
## 2	NA	
## 3	NA	
## 4	NA	
## 5	NA	
## 6	NA	
## diff_Non.Residential.Building.Construction		
## 1	NA	
## 2	NA	
## 3	NA	
## 4	NA	
## 5	NA	
## 6	NA	
## diff_Total.Sales.Tax.Collections.Retail.TX		
## 1	NA	
## 2	NA	
## 3	NA	
## 4	NA	
## 5	NA	
## 6	NA	
## diff_Total.Sales.Tax.Collections.TX diff_Retail.Gasoline.Price.TX		
## 1	NA	NA
## 2	NA	0.068
## 3	NA	0.167
## 4	NA	0.161
## 5	NA	-0.081
## 6	NA	0.013
## diff_Retail.Diesel.Price.TX diff_Nonfarm.Employment.Illinois		
## 1	NA	NA
## 2	0.052	6.6
## 3	0.190	-2.4
## 4	0.078	23.7
## 5	-0.076	-2.1
## 6	0.106	-14.4
## diff_Nonfarm.Employment.Florida diff_Nonfarm.Employment.New.York		
## 1	NA	NA
## 2	20.5	-2.7
## 3	-2.0	-4.3
## 4	47.7	35.0
## 5	36.6	-8.7
## 6	8.6	6.0
## diff_Nonfarm.Employment.Texas diff_Nonfarm.Employment.California		
## 1	NA	NA
## 2	10.8	26.5
## 3	17.6	21.0
## 4	44.3	50.1
## 5	12.0	6.0
## 6	6.9	15.5
## diff_Gross.Value.Crude.Oil.Production diff_Gross.Value.Natural.Gas.Production		
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA

```
## 4          NA          NA
## 5          NA          NA
## 6          NA          NA
##   diff_Motor.Fuel.Taxed.Gasoline diff_Motor.Fuel.Taxed.Diesel
## 1          NA          NA
## 2          NA          NA
## 3          NA          NA
## 4          NA          NA
## 5          NA          NA
## 6          NA          NA
```

Now that we have the differences compared to the previous month as well as the original values, lets look at some correlations. I have chosen the CCI, CPI, Unemployment rate, Home prices, and Energy prices as the target variables to do analysis on.

Correlations

```
library(dplyr)
library(ggplot2)
library(reshape2)

# Assume your dataset is called 'data_diff' (with differenced values)
# Specify your target variables
target_vars <- c( "Unemployment.TX",
                  "Existing.Single.Family.Home.Price.TX",
                  "Retail.Gasoline.Price.TX",
                  "Consumer.Price.Index.TX", "Consumer.Confidence.Index.US")

plot_correlation <- function(target_var, data, threshold = 0) {
  # Filter to remove NA values for the target variable
  filtered_data <- data %>%
    select(all_of(target_var), everything())

  # Calculate the correlation matrix
  cor_matrix <- cor(filtered_data, use = "pairwise.complete.obs")

  # Convert correlation matrix to long format for filtering
  cor_df <- as.data.frame(as.table(cor_matrix))
  colnames(cor_df) <- c("Var1", "Var2", "Correlation")

  # Filter correlations by threshold
  cor_df <- cor_df %>%
    filter(Var1 == target_var & abs(Correlation) > threshold) %>%
    arrange(desc(abs(Correlation)))

  # Create the plot
  ggplot(cor_df, aes(x = Var2, y = Correlation, fill = Correlation)) +
    geom_bar(stat = "identity", position = "dodge") +
    scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0) +
```



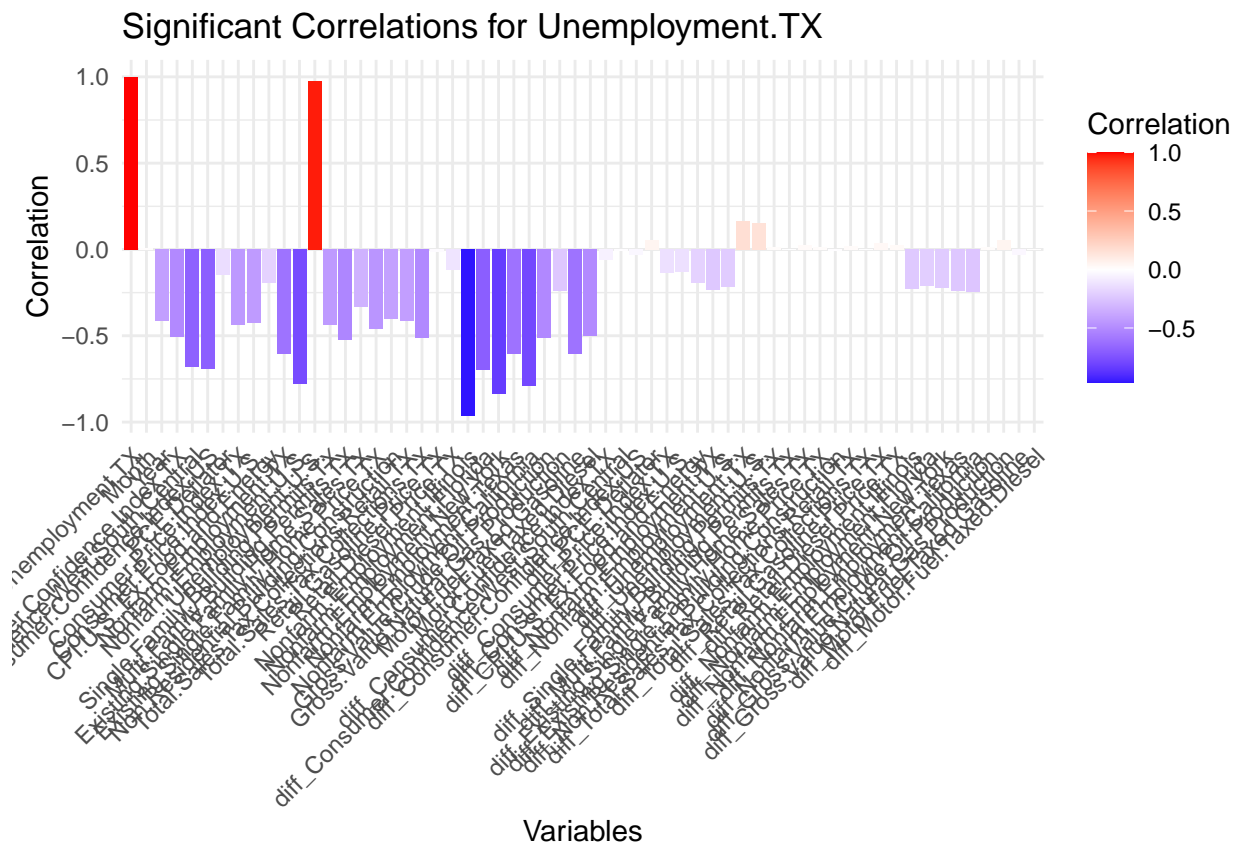
```

theme_minimal() +
ggtitle(paste("Significant Correlations for", target_var)) +
theme(axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotate x-axis labels
xlab("Variables") +
ylab("Correlation")
}

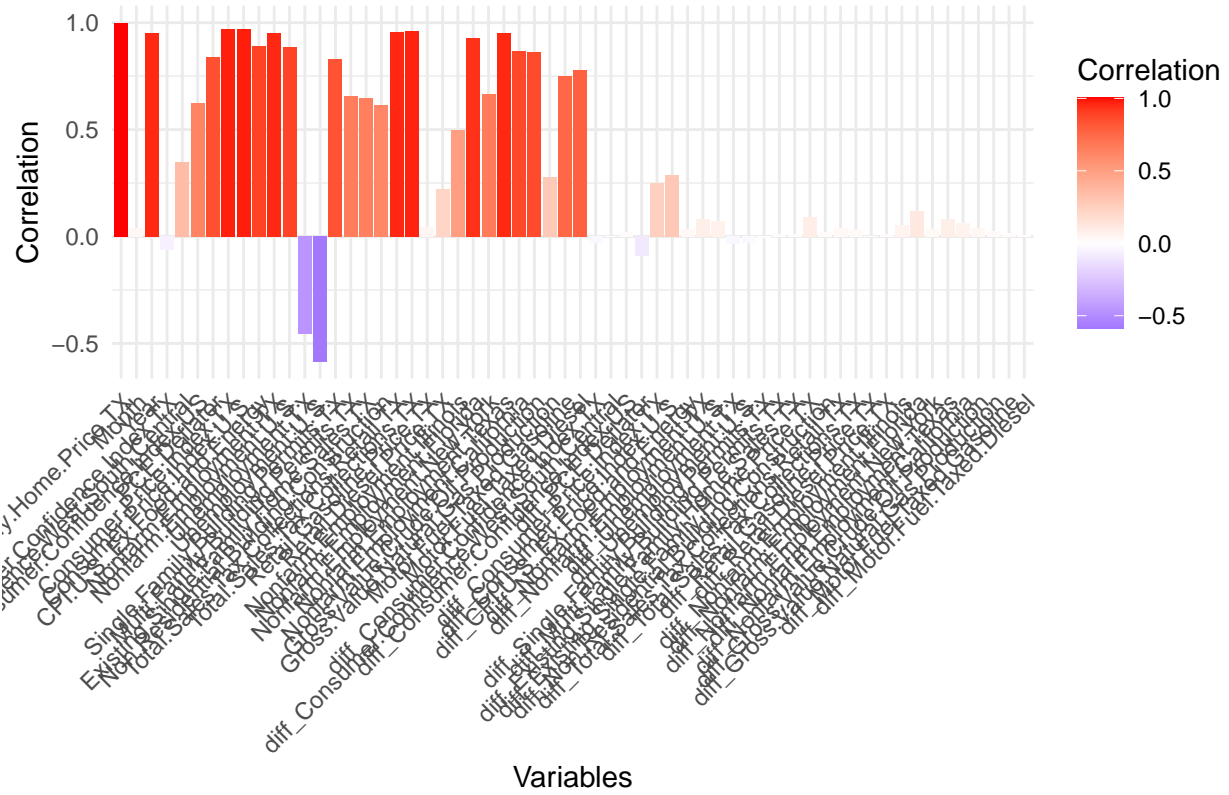
# Generate plots for each target variable
plots <- list()
for (target in target_vars) {
  plots[[target]] <- plot_correlation(target, data_diff)
}

# Print all plots
for (plot_name in names(plots)) {
  print(plots[[plot_name]])
}

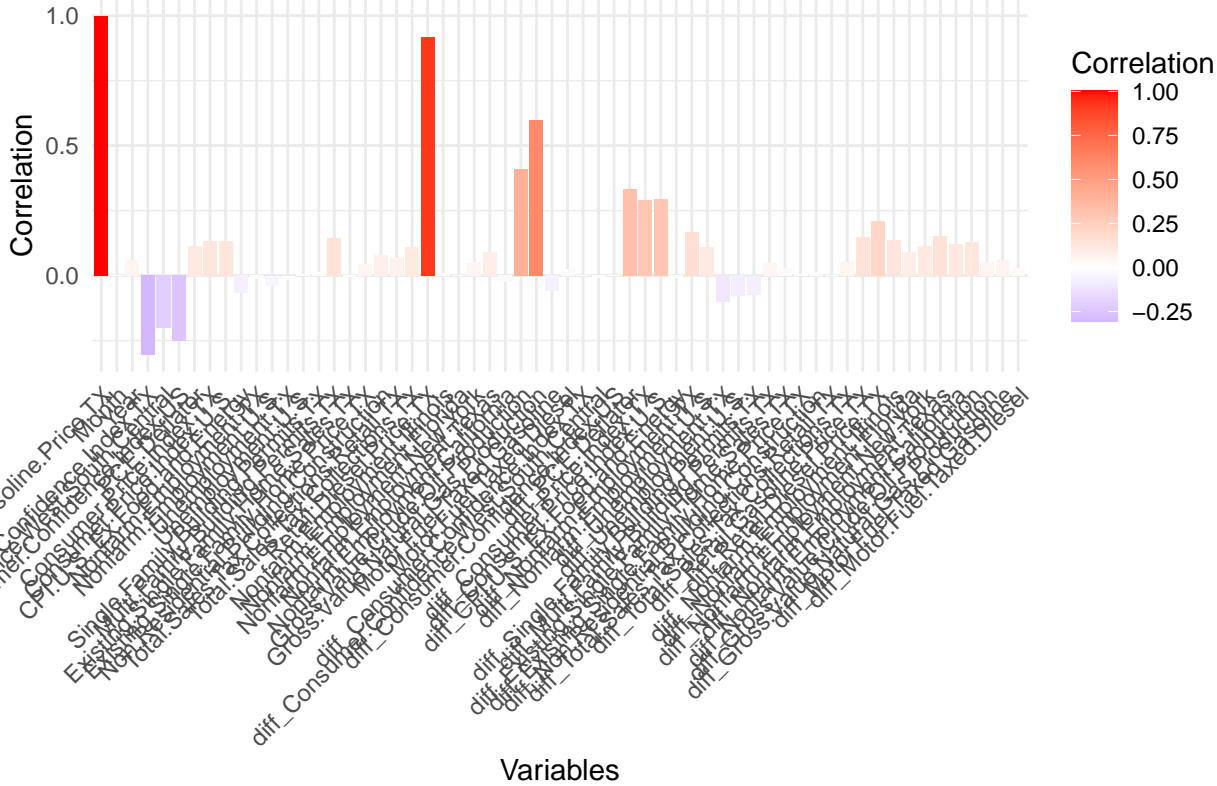
```



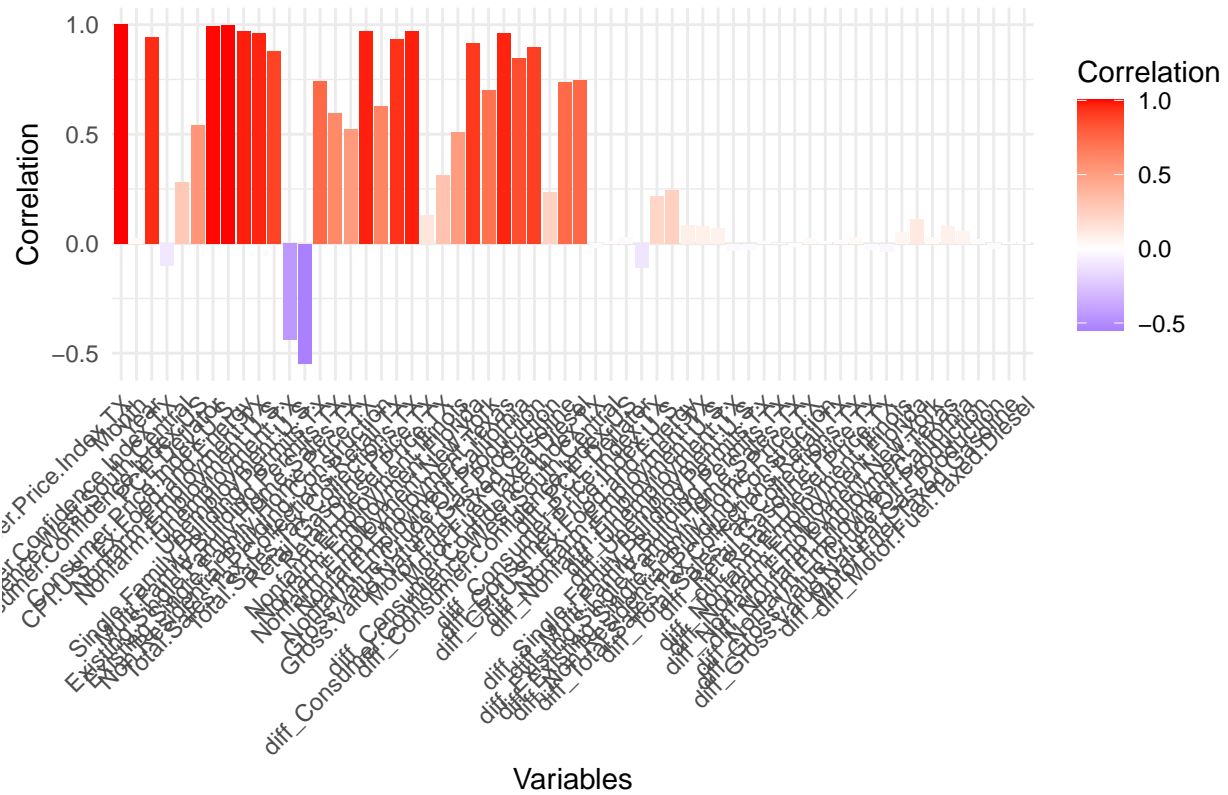
Significant Correlations for Existing.Single.Family.Home.Price.TX

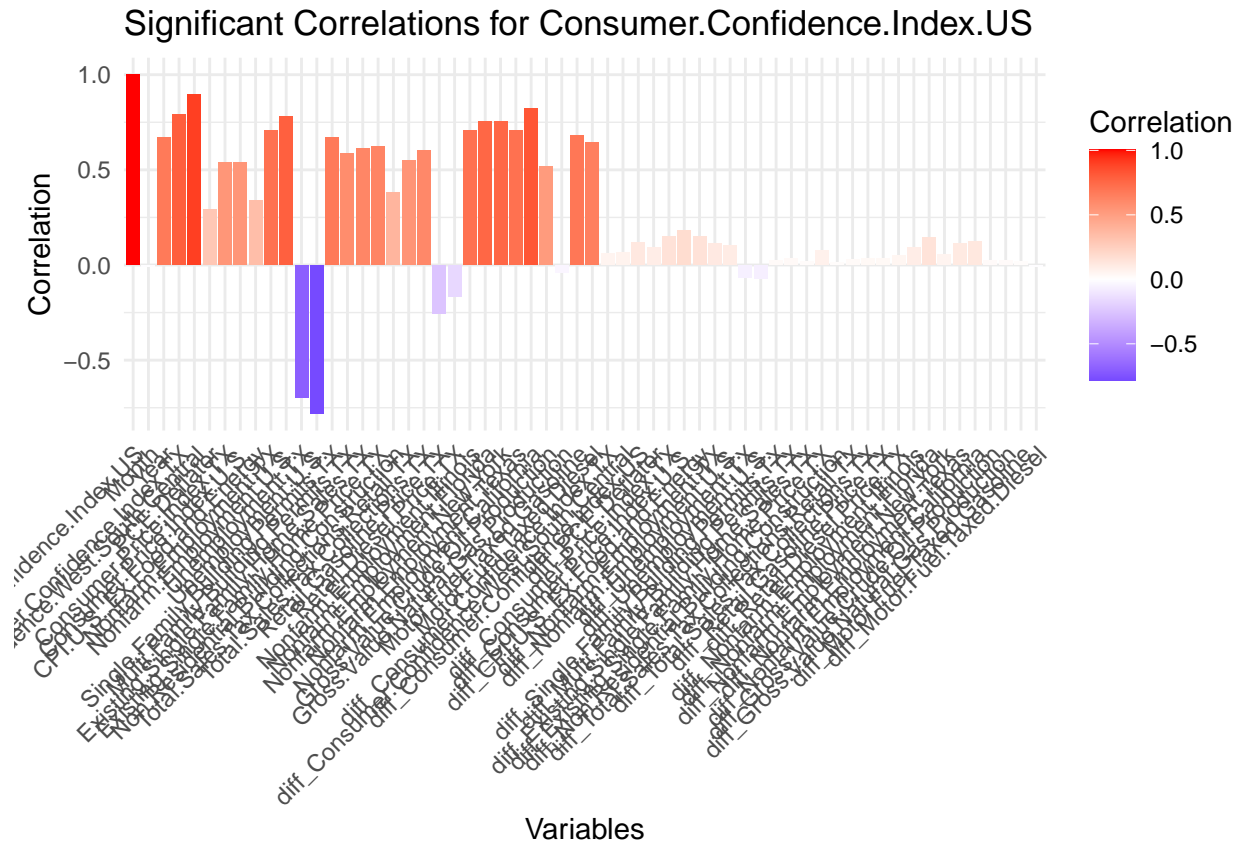


Significant Correlations for Retail.Gasoline.Price.TX



Significant Correlations for Consumer.Price.Index.TX

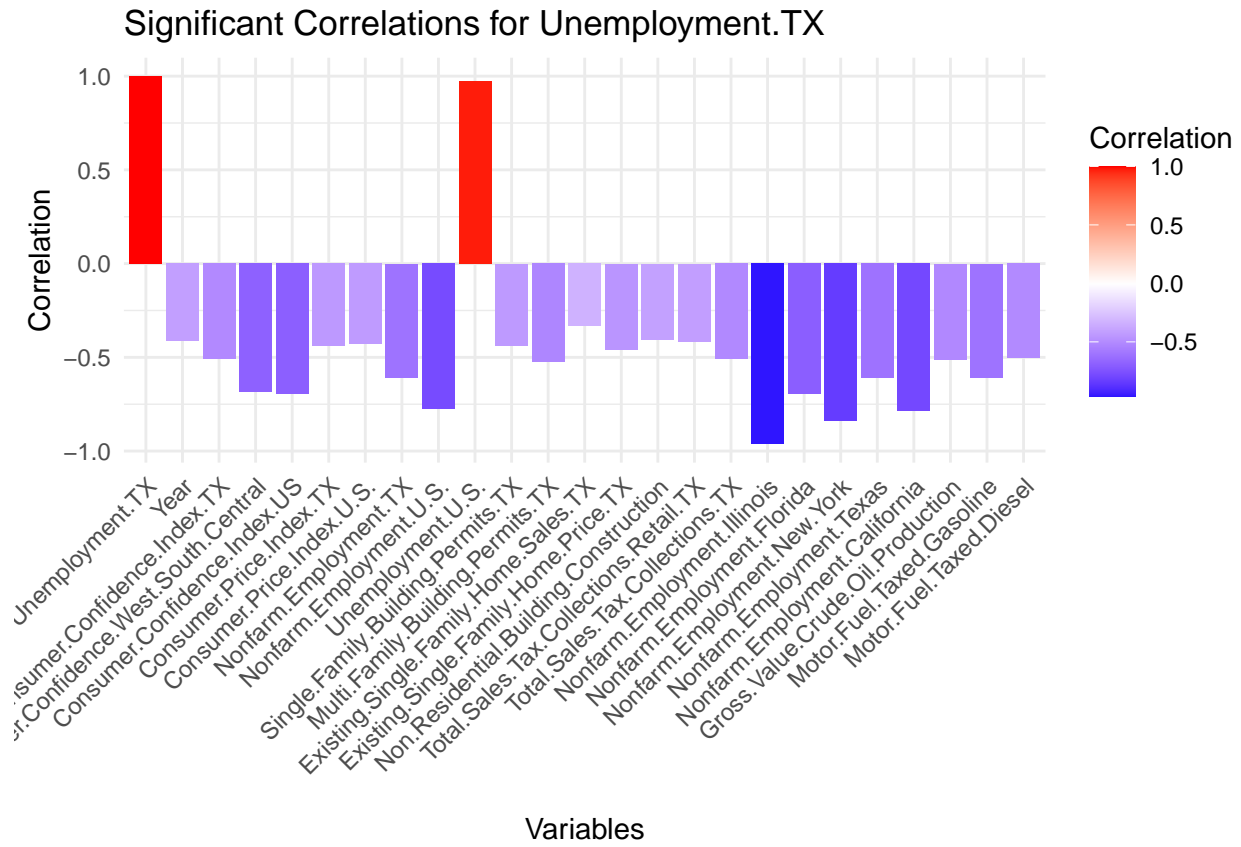


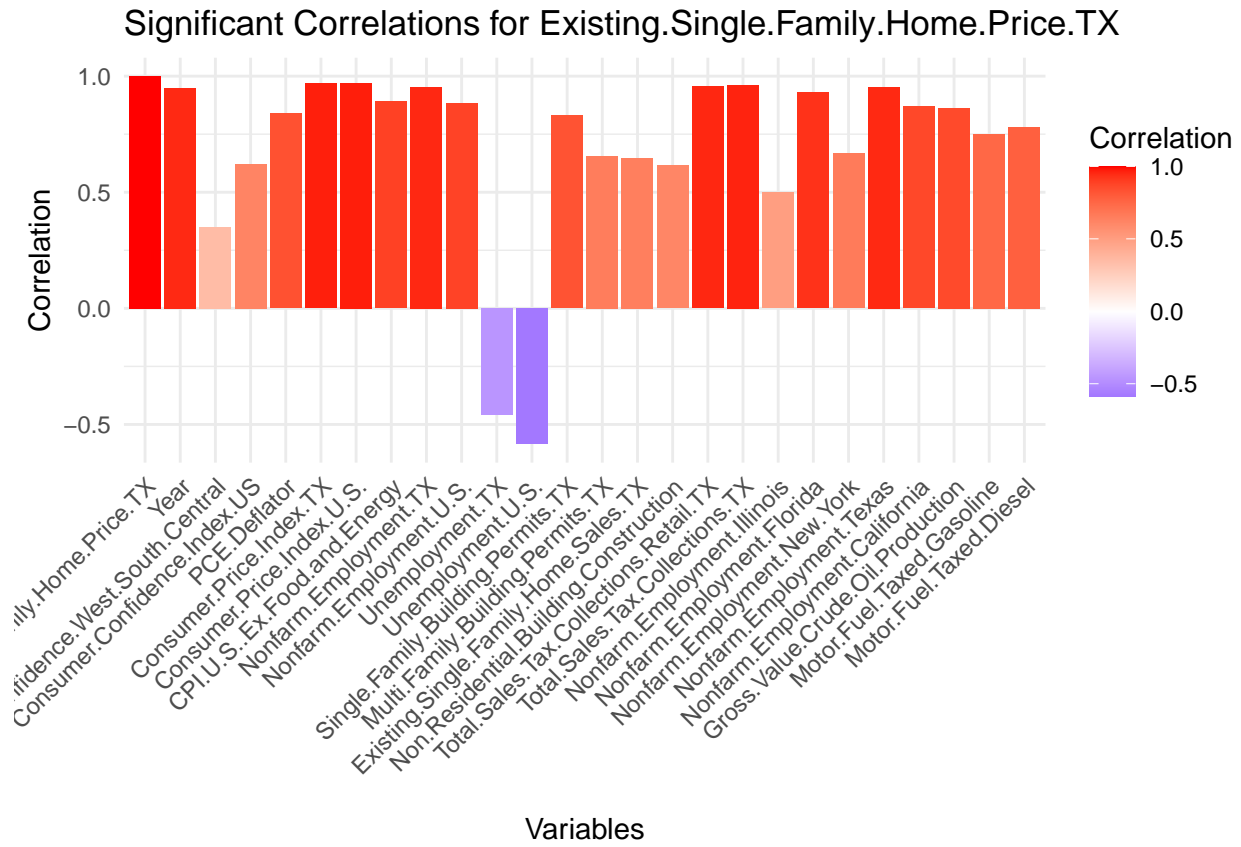


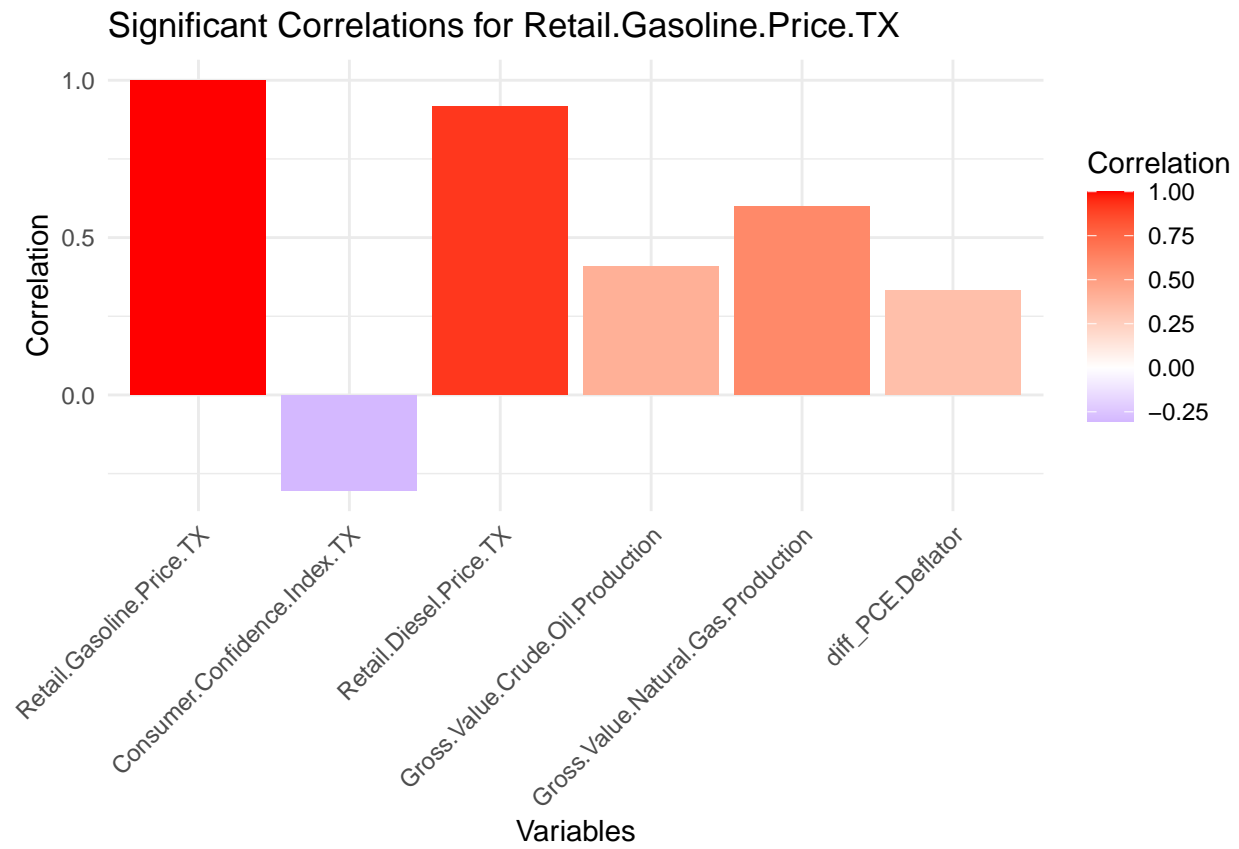
There are clearly a lot of variables here and as a result the graph look quite clustered up. To get a better visualization lets limit it so that variables only variables with $\text{abs}(\text{corr}) > .2$ show up. This helps us better see which variables are most correlated

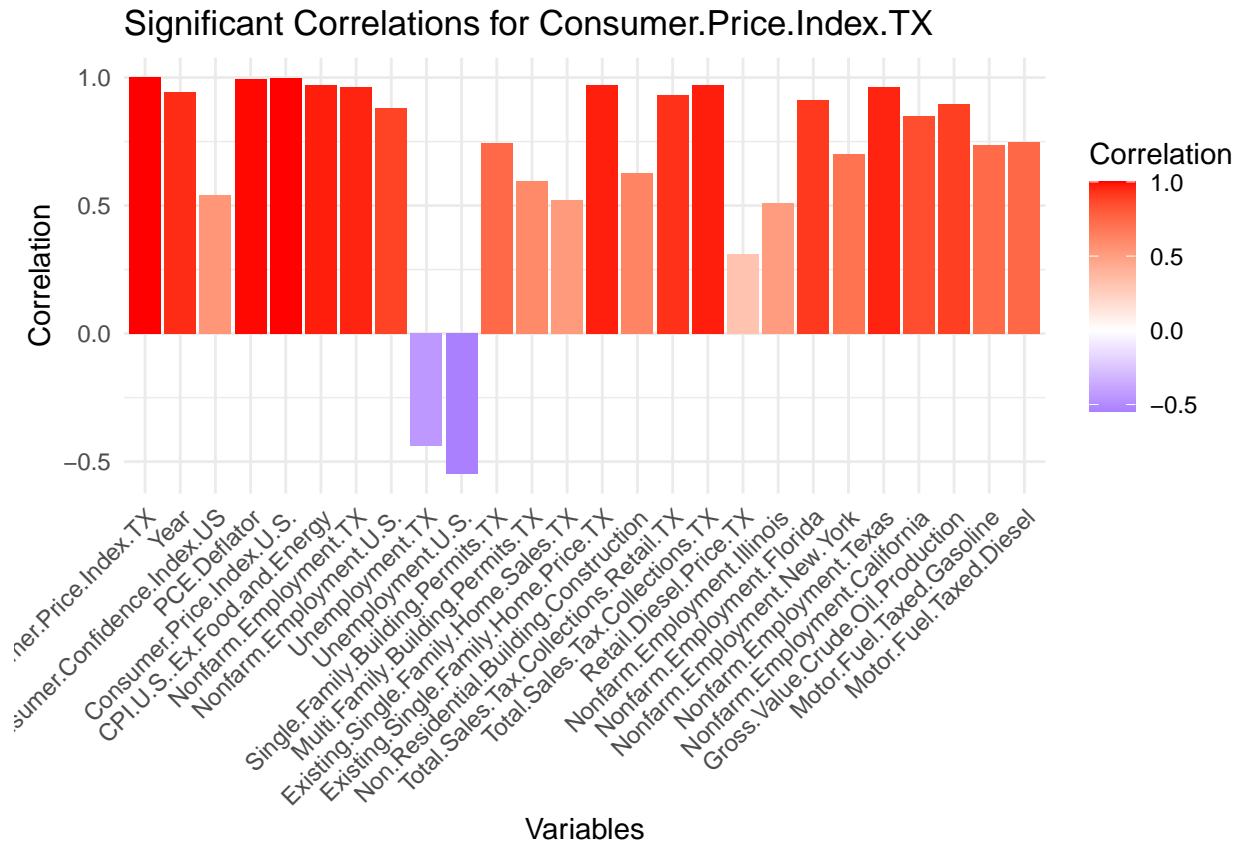
```
plots <- list()
for (target in target_vars) {
  plots[[target]] <- plot_correlation(target, data_diff, .3)
}

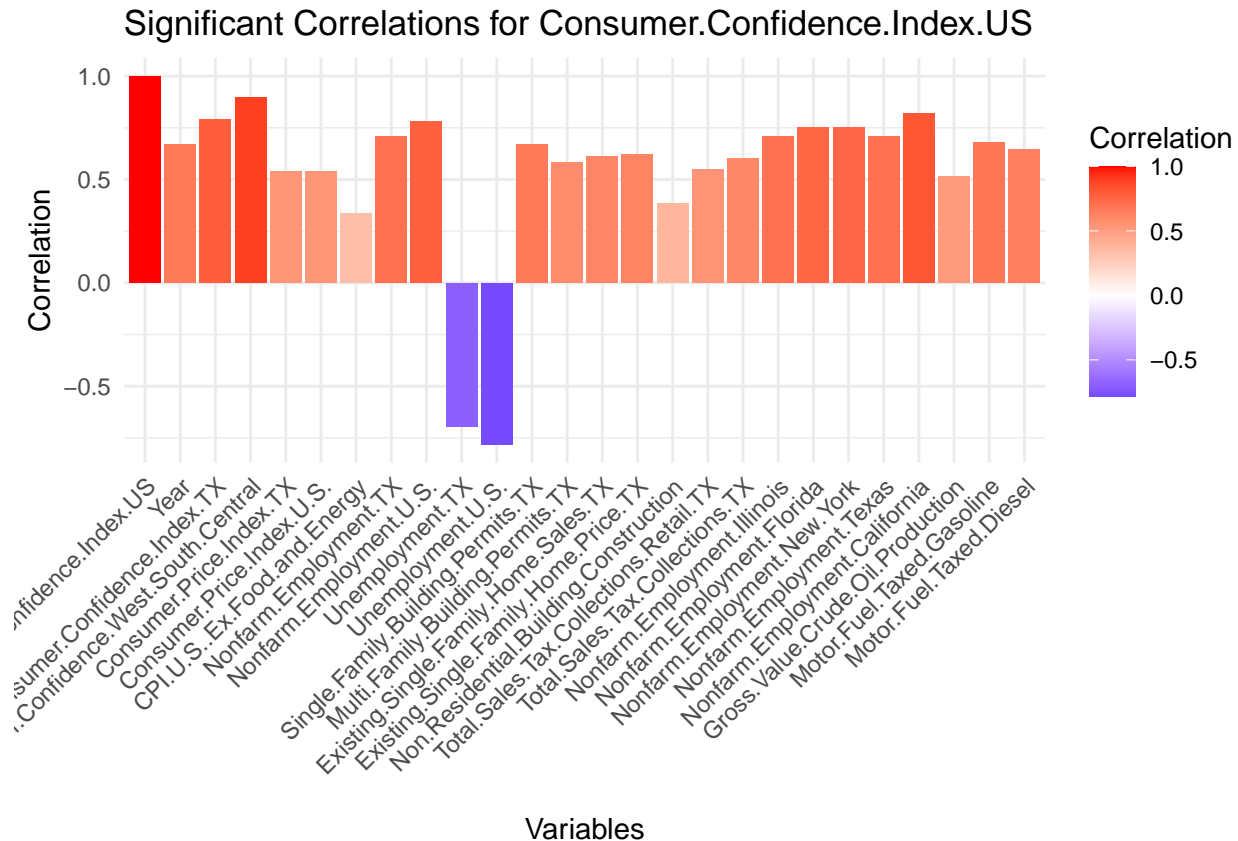
# Print all plots
for (plot_name in names(plots)) {
  print(plots[[plot_name]])
}
```











This is much better for our visualization but maybe lets order it by the magnitude of the correlation to better see which specific ones are important for each target variable

```
plot_correlation <- function(target_var, data, threshold = 0) {
  # Filter to remove NA values for the target variable
  filtered_data <- data %>%
    select(all_of(target_var), everything())

  # Calculate the correlation matrix
  cor_matrix <- cor(filtered_data, use = "pairwise.complete.obs")

  # Convert correlation matrix to long format for filtering
  cor_df <- as.data.frame(as.table(cor_matrix))
  colnames(cor_df) <- c("Var1", "Var2", "Correlation")

  # Filter correlations by threshold
  cor_df <- cor_df %>%
    filter(Var1 == target_var & abs(Correlation) > threshold & Var2 != target_var) %>%
    arrange(desc(Correlation))

  # Reorder Var2 for plotting
  cor_df$Var2 <- factor(cor_df$Var2, levels = cor_df$Var2)

  # Create the plot
  ggplot(cor_df, aes(x = Var2, y = Correlation, fill = Correlation)) +
    geom_bar(stat = "identity", position = "dodge") +
```

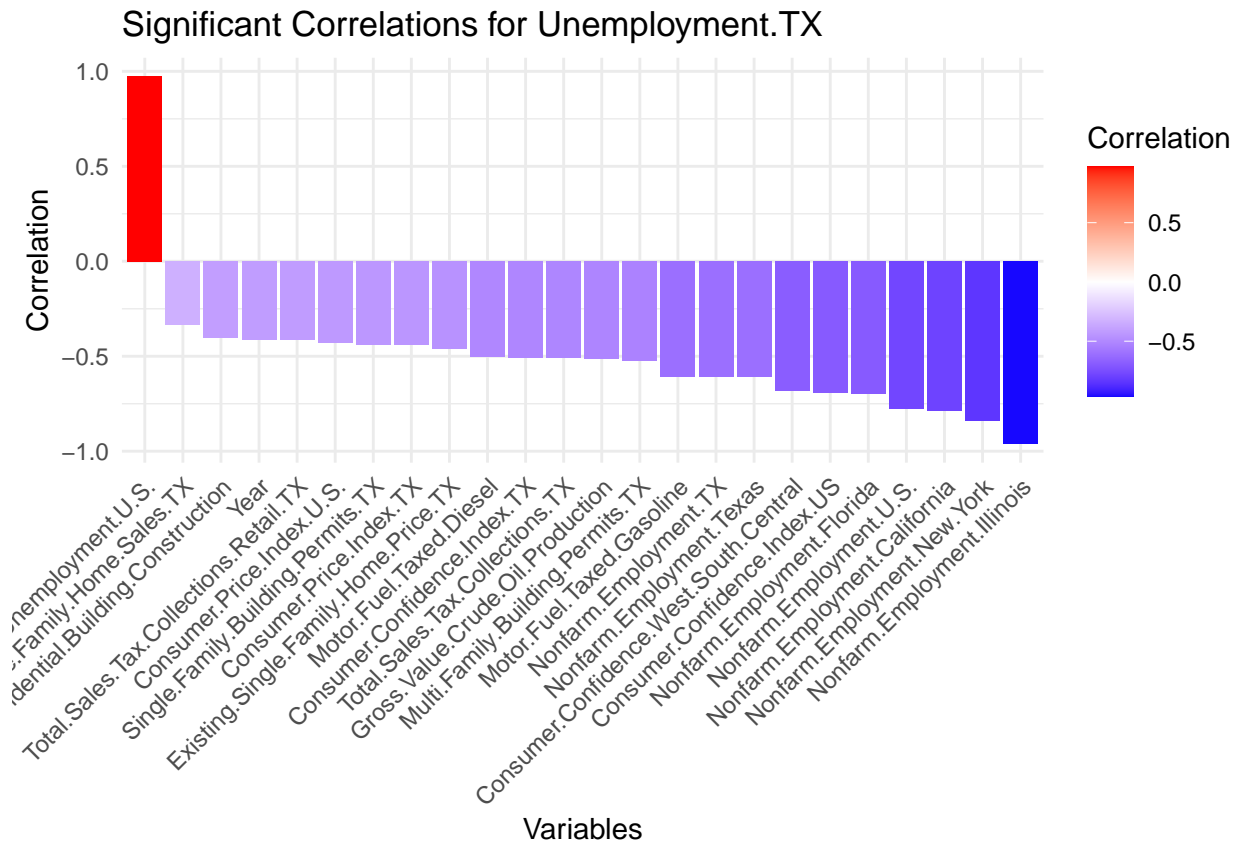
```

scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0) +
theme_minimal() +
ggtitle(paste("Significant Correlations for", target_var)) +
theme(axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotate x-axis labels
xlab("Variables") +
ylab("Correlation")
}

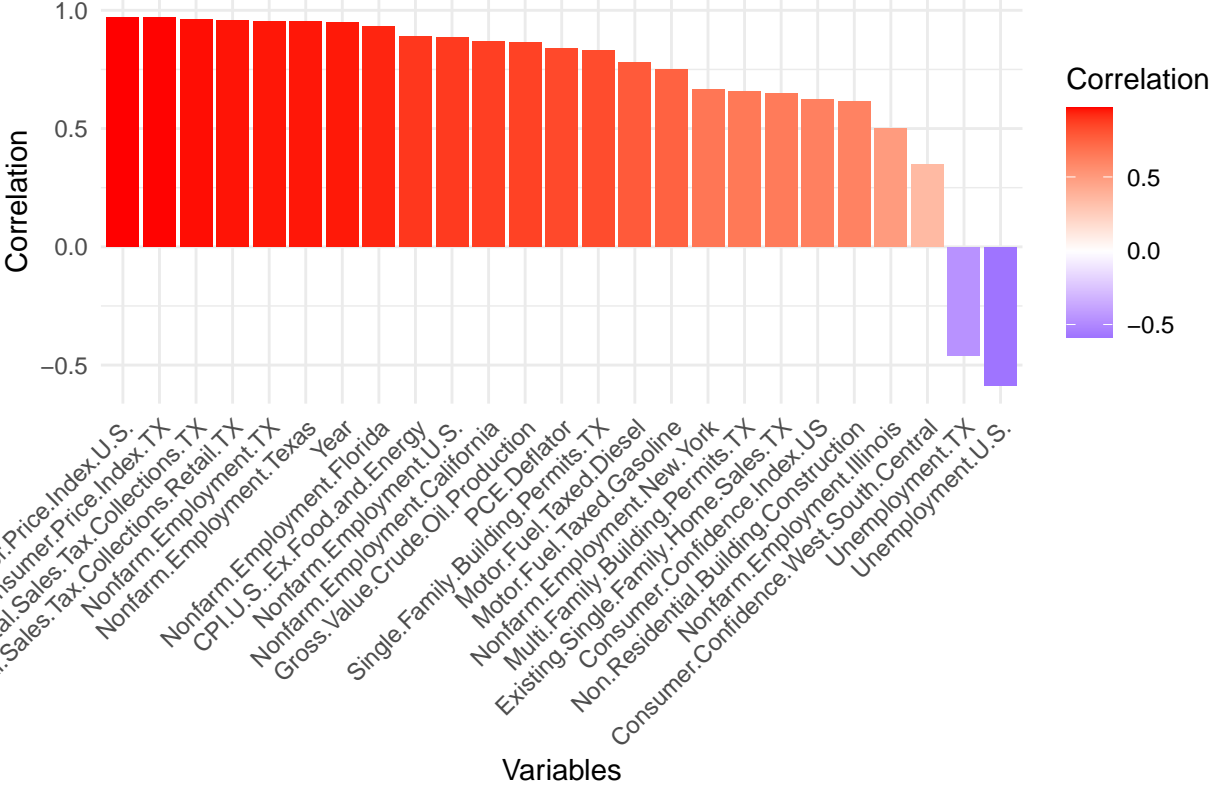
# Generate plots for each target variable
plots <- list()
for (target in target_vars) {
  plots[[target]] <- plot_correlation(target, all_econ_data, .3)
}

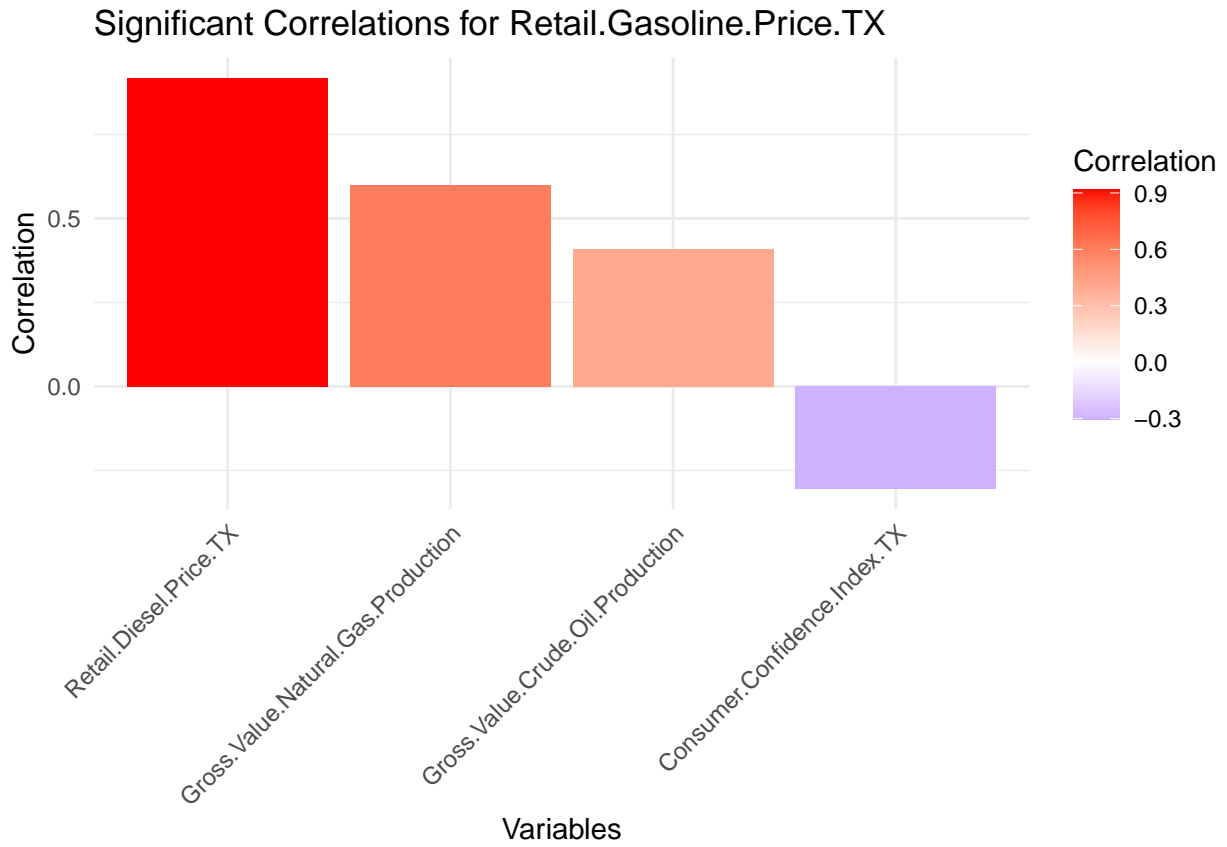
# Print all plots
for (plot_name in names(plots)) {
  print(plots[[plot_name]])
}

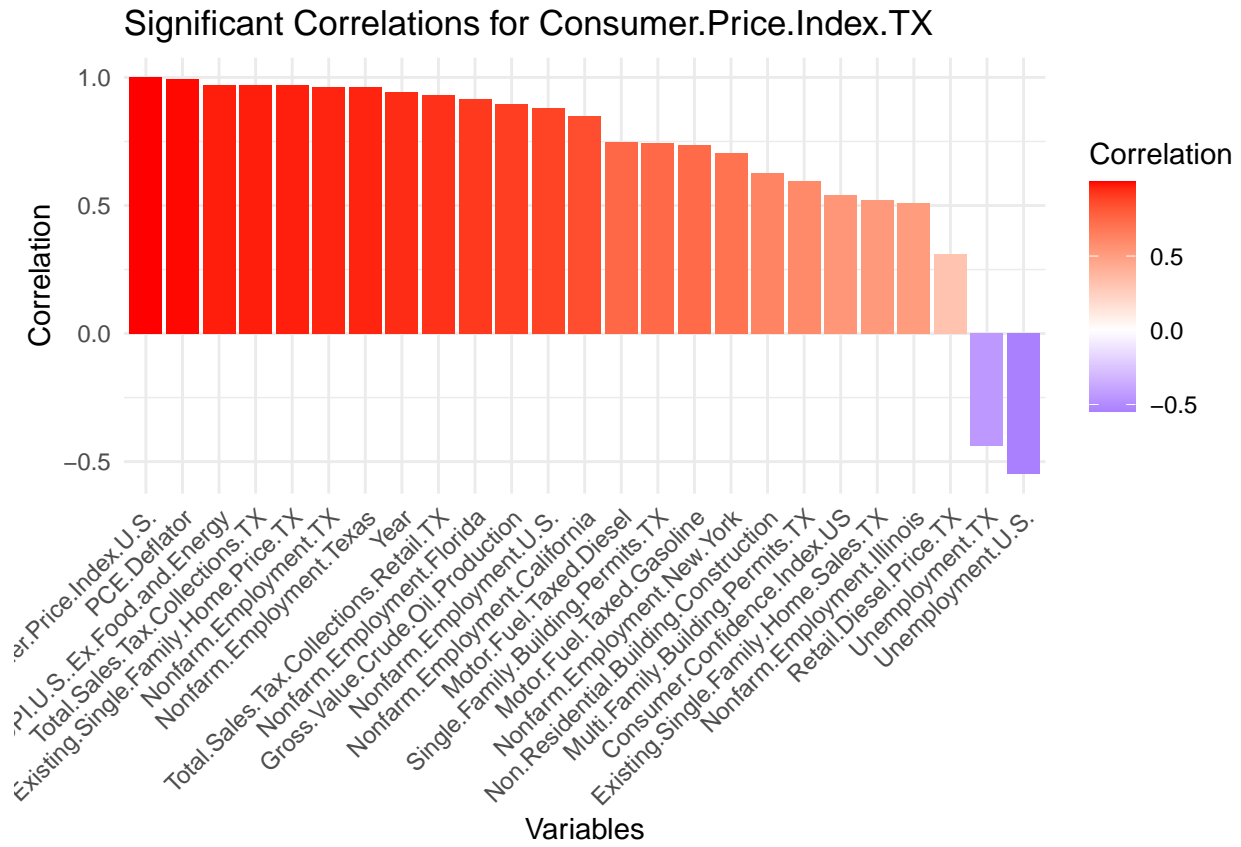
```

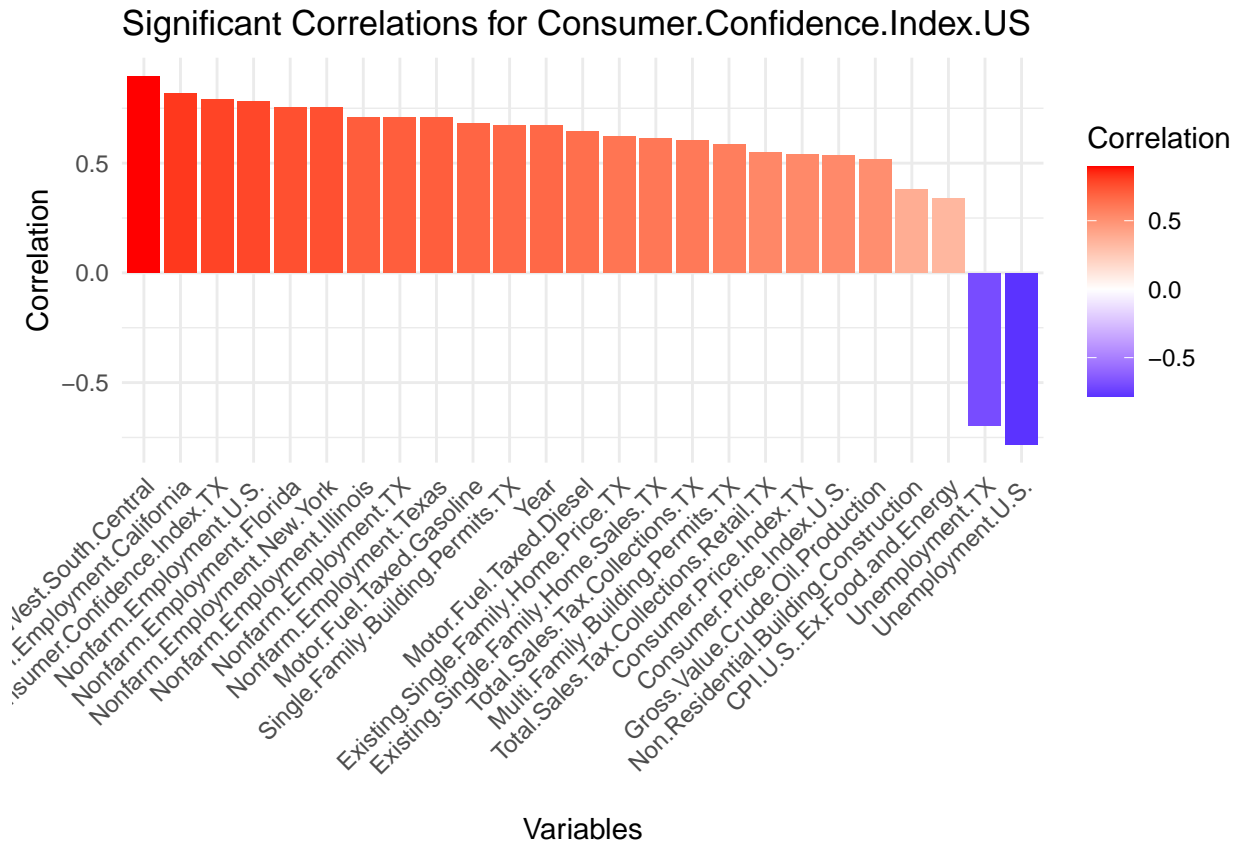


Significant Correlations for Existing.Single.Family.Home.Price.TX









Correlation with Year ##### I clearly overlooked the Year variable in these correlations and did not take it out but it appears as a significant correlator to 3 out of 5 of the target variables. So out of curiosity lets graph each of them as compred to their average value for that year.

```
library(tidyr)

##
## Attaching package: 'tidyr'

## The following object is masked from 'package:reshape2':
##
## smiths

yearly_avg <- all_econ_data %>%
  group_by(Year) %>%
  summarise(across(all_of(target_vars), mean, na.rm = TRUE)) %>%
  pivot_longer(-Year, names_to = "Variable", values_to = "Average")

## Warning: There was 1 warning in 'summarise()'.
## i In argument: 'across(all_of(target_vars), mean, na.rm = TRUE)'.
## i In group 1: 'Year = 2005'.
## Caused by warning:
## ! The '...' argument of 'across()' is deprecated as of dplyr 1.1.0.
## Supply arguments directly to '.fns' through an anonymous function instead.
##
```

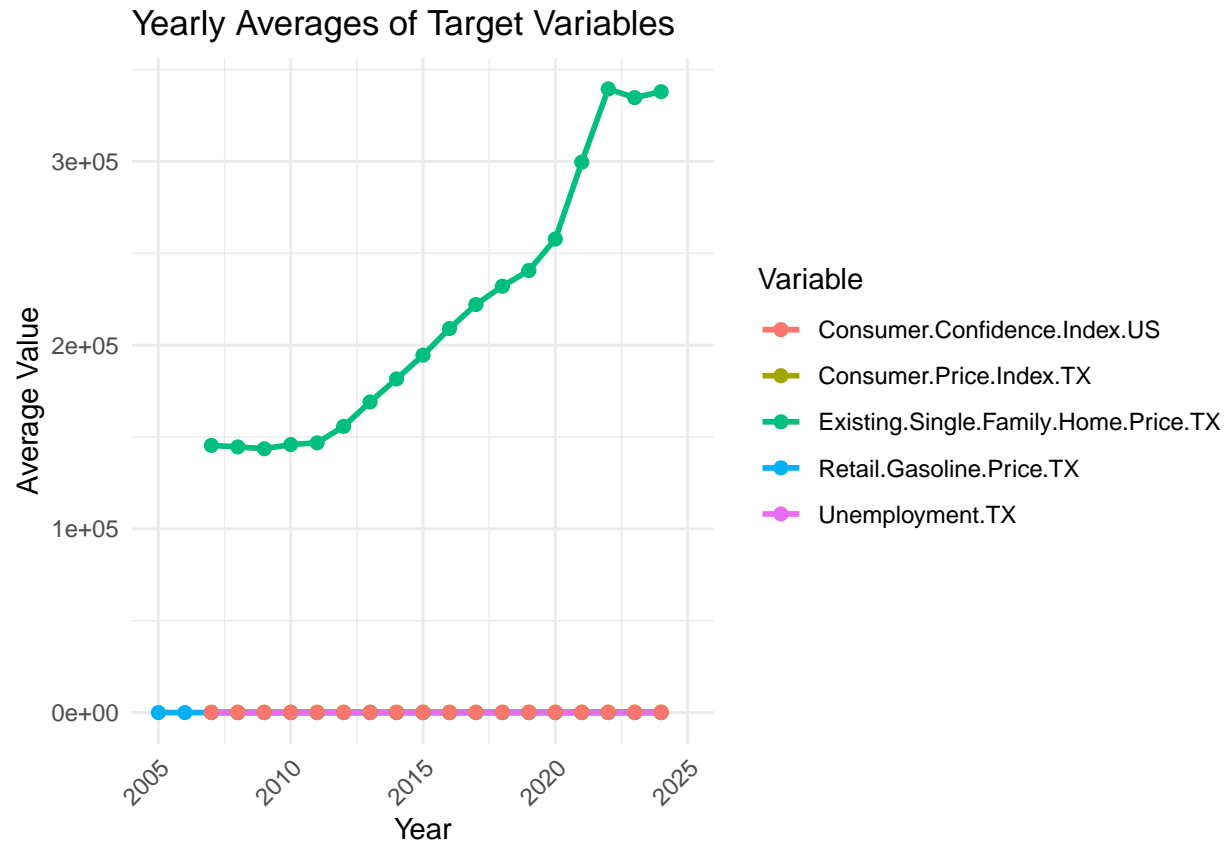
```
## # Previously
## across(a:b, mean, na.rm = TRUE)
##
## # Now
## across(a:b, \(x) mean(x, na.rm = TRUE))
```

```
# Plot yearly averages for each target variable
ggplot(yearly_avg, aes(x = Year, y = Average, color = Variable, group = Variable)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  theme_minimal() +
  labs(
    title = "Yearly Averages of Target Variables",
    x = "Year",
    y = "Average Value",
    color = "Variable"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
## Warning: Removed 13 rows containing missing values ('geom_line()').
```

```
## Warning: Removed 13 rows containing missing values ('geom_point()').
```

Although this might make it seem like only the house prices have a true correlation, I believe this is because house prices are huge compared to the other variables. So how about let's standardize the values in each column so that we can better see this change through the years

```
# Standardize the target variables
all_econ_data_standardized <- all_econ_data %>%
  mutate(across(all_of(target_vars), ~ (. - mean(., na.rm = TRUE)) / sd(., na.rm = TRUE)),
    .names = "std_{col}")

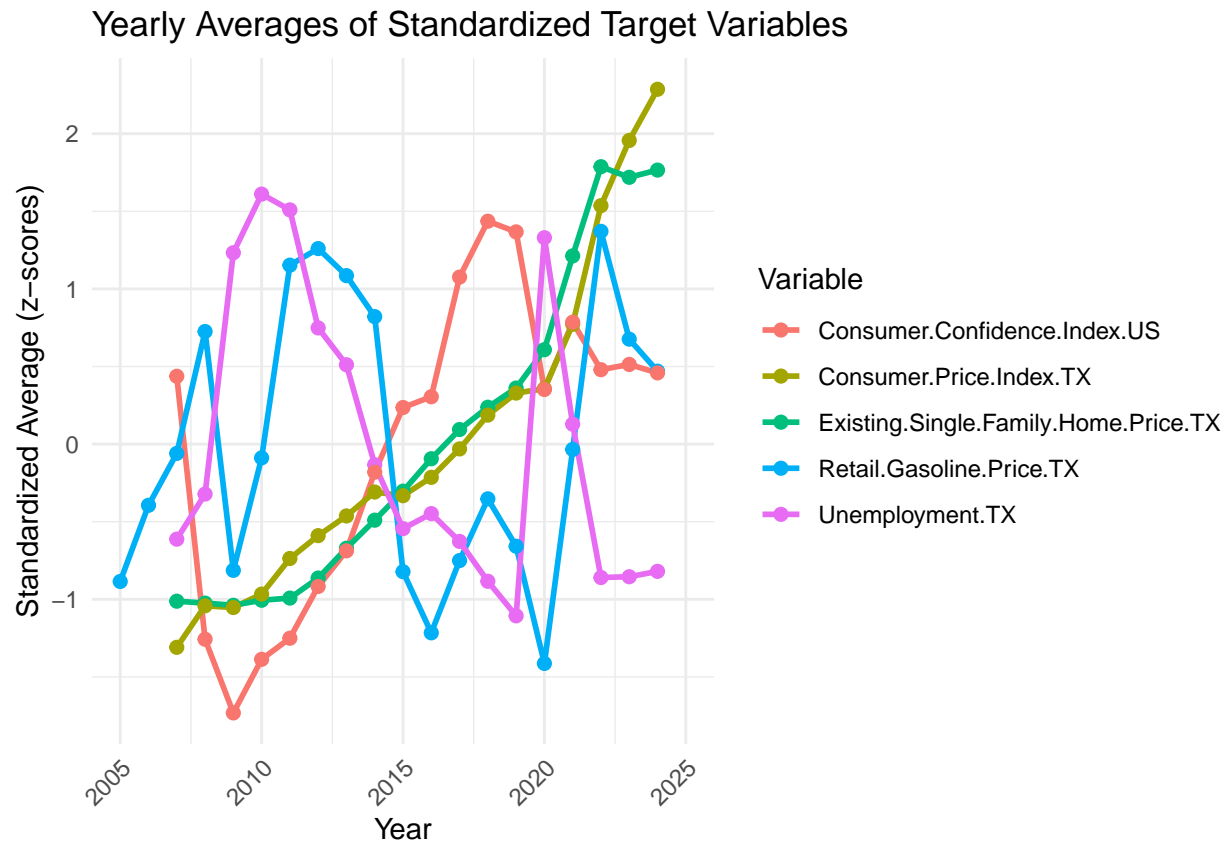
# Calculate yearly averages for standardized variables
yearly_avg_std <- all_econ_data_standardized %>%
  group_by(Year) %>%
  summarise(across(starts_with("std_"), mean, na.rm = TRUE)) %>%
  pivot_longer(-Year, names_to = "Variable", values_to = "Average") %>%
  mutate(Variable = gsub("std_", "", Variable)) # Clean variable names for plot

# Plot yearly averages for standardized variables
ggplot(yearly_avg_std, aes(x = Year, y = Average, color = Variable, group = Variable)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  theme_minimal() +
  labs(
    title = "Yearly Averages of Standardized Target Variables",
    x = "Year",
    y = "Standardized Average (z-scores)",
    color = "Variable"
  ) +
```

```
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
## Warning: Removed 13 rows containing missing values ('geom_line()').
```

```
## Warning: Removed 13 rows containing missing values ('geom_point()').
```



Well it does make sense that house prices and the CPI would increase with the years and it can also be seen here on the graph. But it still doesn't make sense why unemployment a correlation because it seems like most of them fluctuate through the years. And it intuitively makes sense because the unemployment rate and CCI should fluctuate as they fluctuate with the economy, but I guess gasoline prices also fluctuates and from what we see on the graph it moves quite similarly to the unemployment rate.

PCA

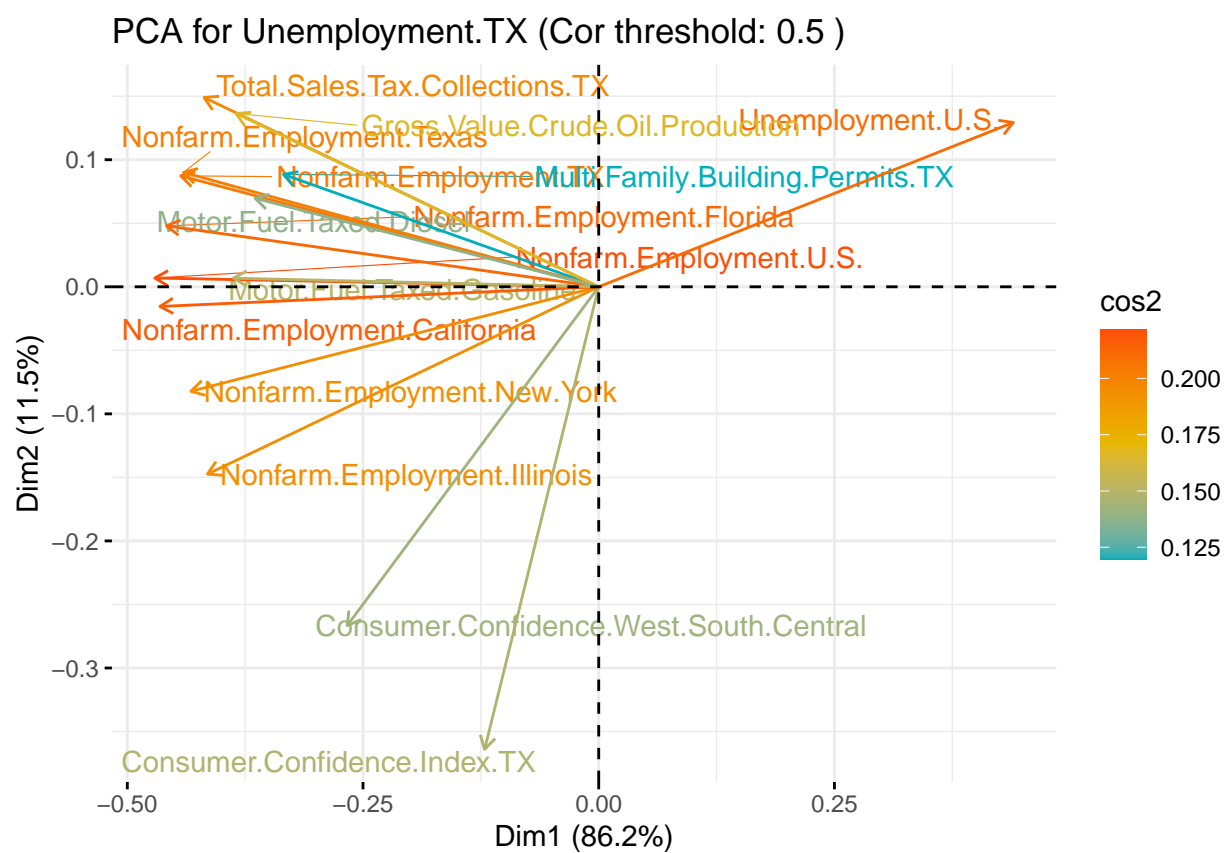
```
library(factoextra)
```

Another way to look at correlations is to perform PCA and extract the variables that seem to contribute the most to the variance in the target variables hence the ones that have the largest impact on the target variable. To make things a bit simpler I will only visualize the first 2 PCs but these 2 should be able to explain majority of the variance in the dataset. Also just like before in order to avoid a cluster of all the features I will only consider features that have a correlation higher than a specified threshold starting with threshold=0.3

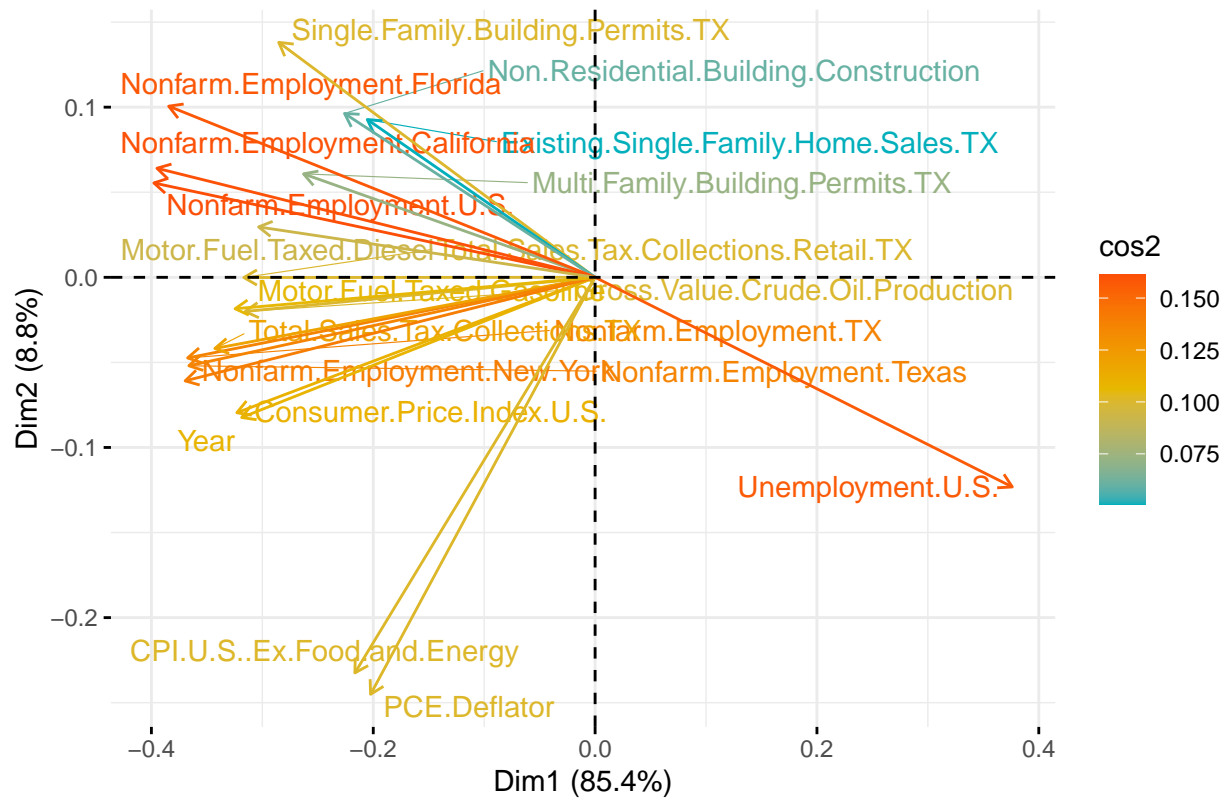
Welcome! Want to learn more? See two factoextra-related books at <https://goo.gl/ve3WBa>

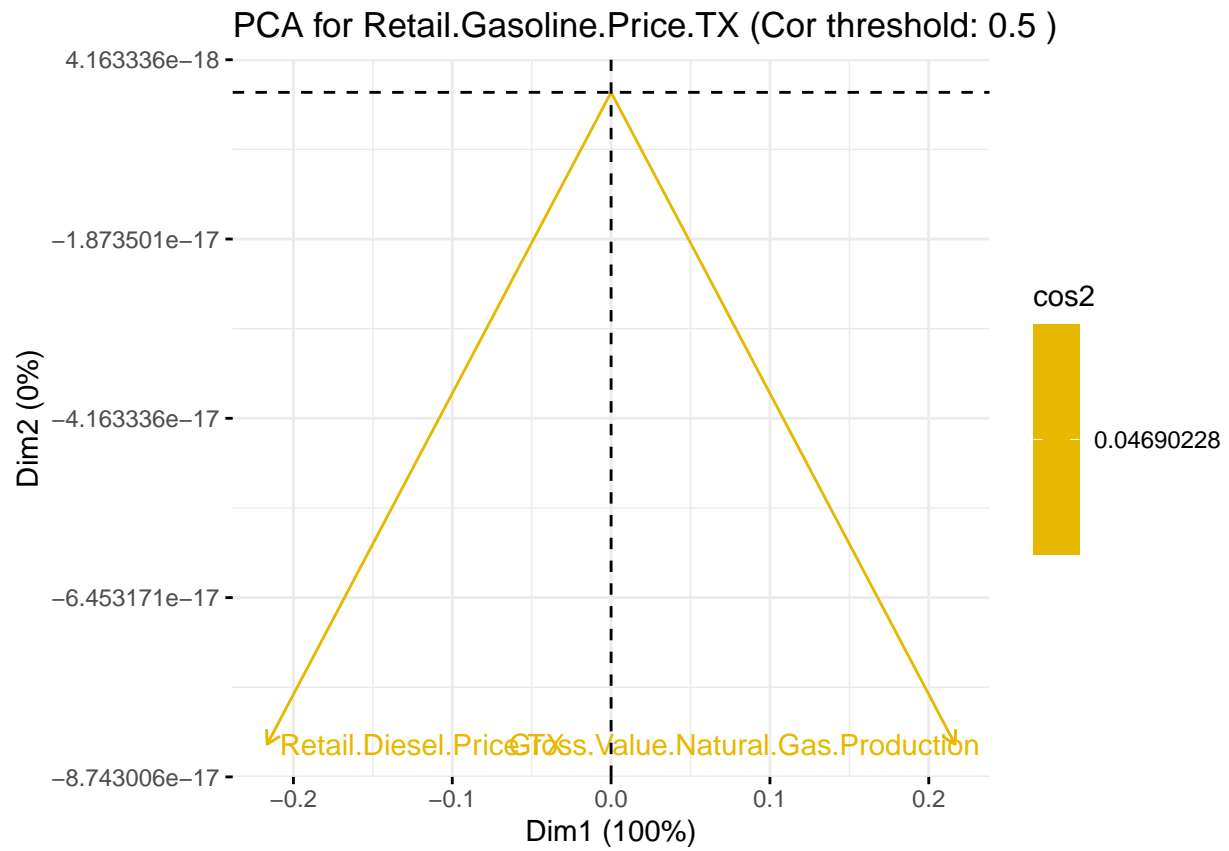
```
perform_pca_with_inline_visualization <- function(data, target_vars, cor_threshold = 0.5) {  
  # Initialize a list to store PCA results for each target variable  
  pca_results <- list()  
  
  for (target_var in target_vars) {  
  
    # Subset data for pairwise complete cases with the target variable  
    target_data <- data[, c(target_var, setdiff(colnames(data), target_vars))]  
  
    # Compute the correlation matrix using pairwise.complete.obs  
    cor_matrix <- cor(target_data, use = "pairwise.complete.obs")  
  
    # Filter variables based on correlation threshold with the target variable  
    target_correlations <- cor_matrix[, target_var, drop = FALSE] # Correlations with the target variable  
    vars_to_include <- rownames(target_correlations)[abs(target_correlations[, 1]) >= cor_threshold]  
    vars_to_include <- setdiff(vars_to_include, target_var)  
  
    if (length(vars_to_include) < 2) {  
      warning(paste("Not enough variables meet the correlation threshold for", target_var, ". Skipping to next"))  
      next  
    }  
  
    # Subset the correlation matrix to include only selected variables  
    filtered_cor_matrix <- cor_matrix[vars_to_include, vars_to_include]  
  
    # Perform PCA on the filtered correlation matrix  
    pca_result <- prcomp(filtered_cor_matrix, scale. = FALSE)  
  
    # Store the PCA result in the list  
    pca_results[[target_var]] <- list(  
      pca = pca_result,  
      explained_variance = summary(pca_result)$importance[2, ] # Proportion of variance explained  
    )  
  
    # Visualization  
    plot_title <- paste("PCA for", target_var, "(Cor threshold:", cor_threshold, ")")  
    pca_plot <- fviz_pca_var(  
      pca_result,  
      title = plot_title,  
      repel = TRUE,  
      col.var = "cos2", # Coloring by contribution  
      gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07")  
    )  
  
    # Print the plot inline  
    print(pca_plot)  
  }  
  
  return(pca_results)  
}
```

```
# Example usage with a correlation threshold of 0.6
pca_results <- perform_pca_with_inline_visualization(all_econ_data, target_vars, cor_threshold = 0.5)
```

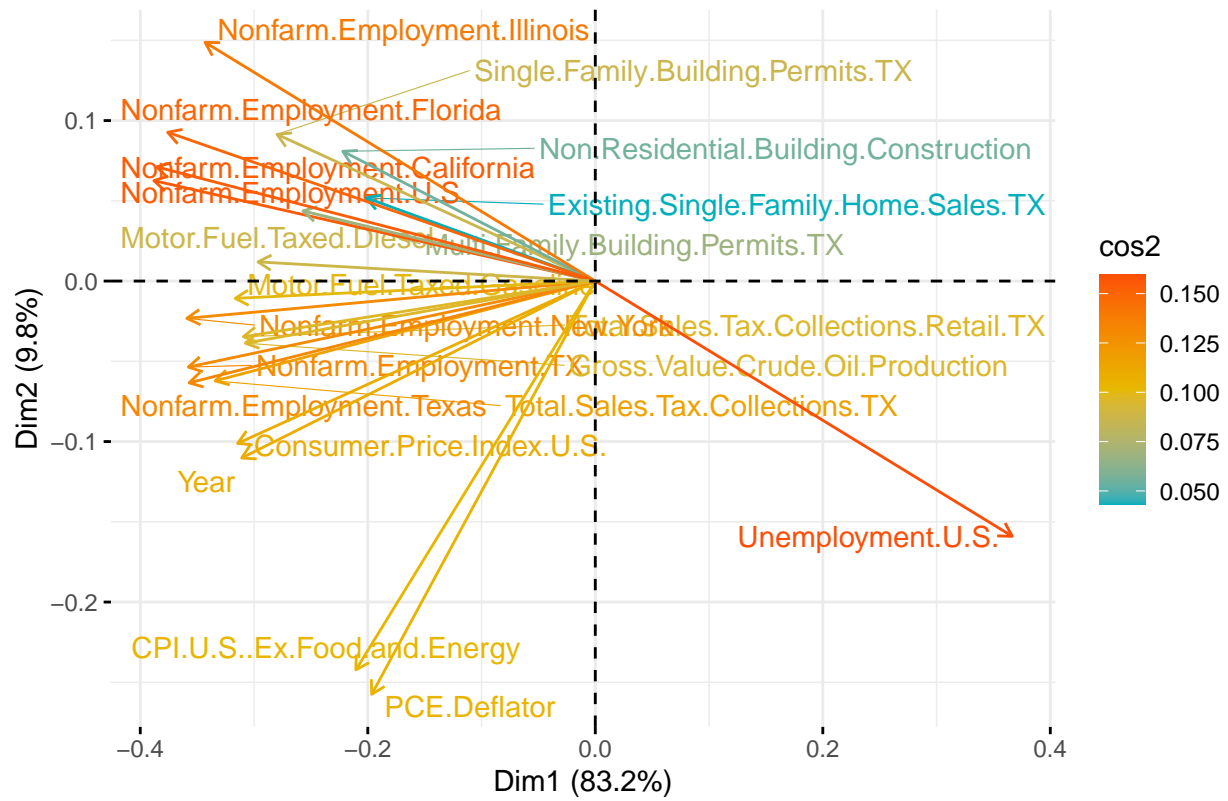


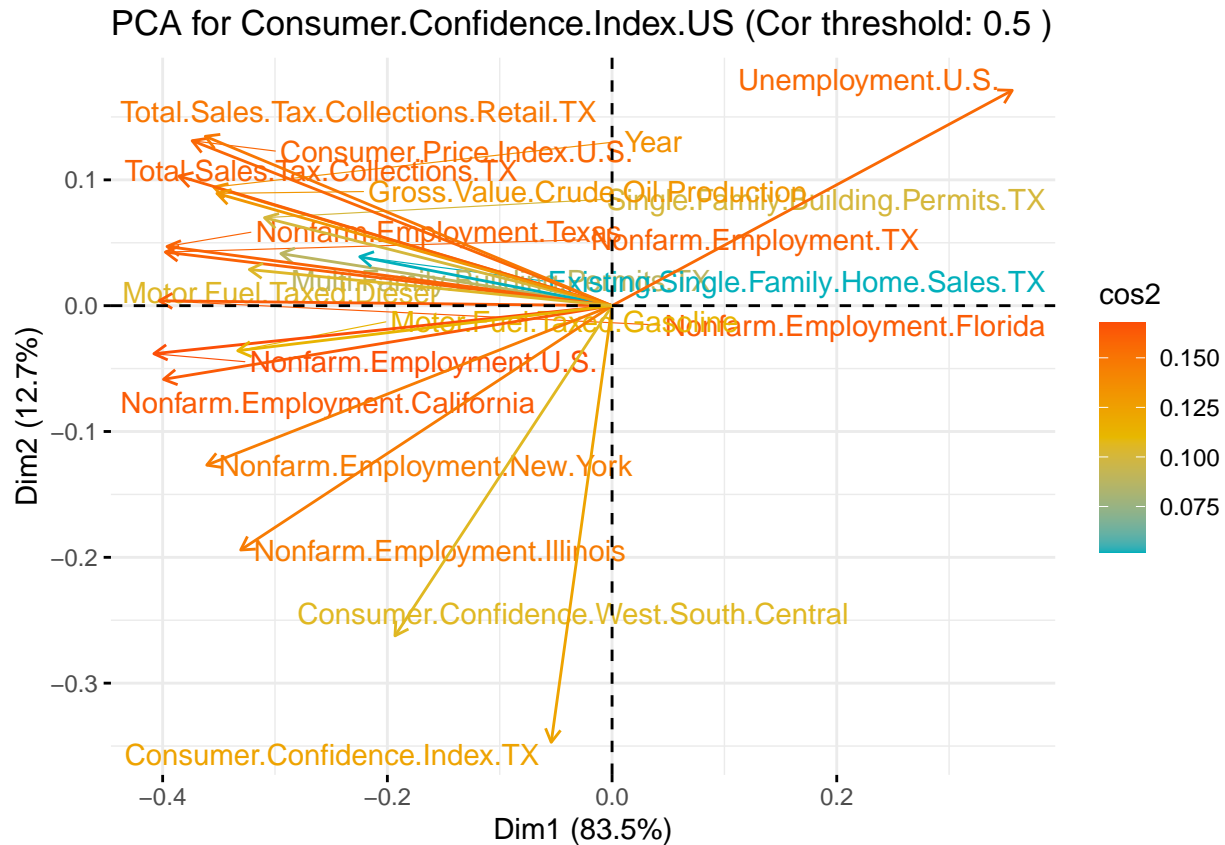
PCA for Existing.Single.Family.Home.Price.TX (Cor threshold: 0.5)





PCA for Consumer.Price.Index.TX (Cor threshold: 0.5)





Let's analyze CCI Index's Principal Component(PC) graph as it has many features still left on the graph whilst not being as clustered as the other ones. We see 2 principal components on the graph with PC1 being on the Y axis and PC2 being on the x axis. For CCI PC1 explains 83.5% of the variance and PC2 explains 12.7% as is shown by the x and y axis labels on the graphs meaning together they explain $76.6+12.4=96.2\%$ of the variance seen in the CCI for Texas. We can also see that the Unemployment rate in US contributes positively to both of these PCs' values as it appears in the first quadrant meaning it has a positive component for both PCs. Similarly all the features appearing in the 3rd quadrant of this PCA graph all contribute negatively to both of the PCs' values. On this graph we can also see that features that have high positive correlation with each other tend to have smaller angles(close to 0) between their component lines and features with high negative correlation tend to point in the opposite directions(180 degrees between them). For example all the Nonfarm Employment vectors in quadrant 2 and 3 have a very small angle between them and it makes sense that they would all be highly correlated with each other.