A study on GDP of Republic of Moldova and its correlation with Google Trends

Initial visualization of the GDP dataset

The question of interest is the the relationship between GDP in Republic of Moldova and it's correlation with the Google searches in different categories. The data for the quarterly GDP of Republic of Moldova are from the National Bureau of Statistics and the Google searches information are from Google Trends.

In our notation, the GDP data points are GDP_t with $t \in \{2013.25, 2013.5, 2013.75, \ldots, \ldots 2023\}$. That is, we know the nominal GDP for every quarter from 2013 to 2022 inclusive, a total of 40 data points.

Below you can see a plot of the Republic of Moldova GDP in every quarter, from 2013 to 2023.

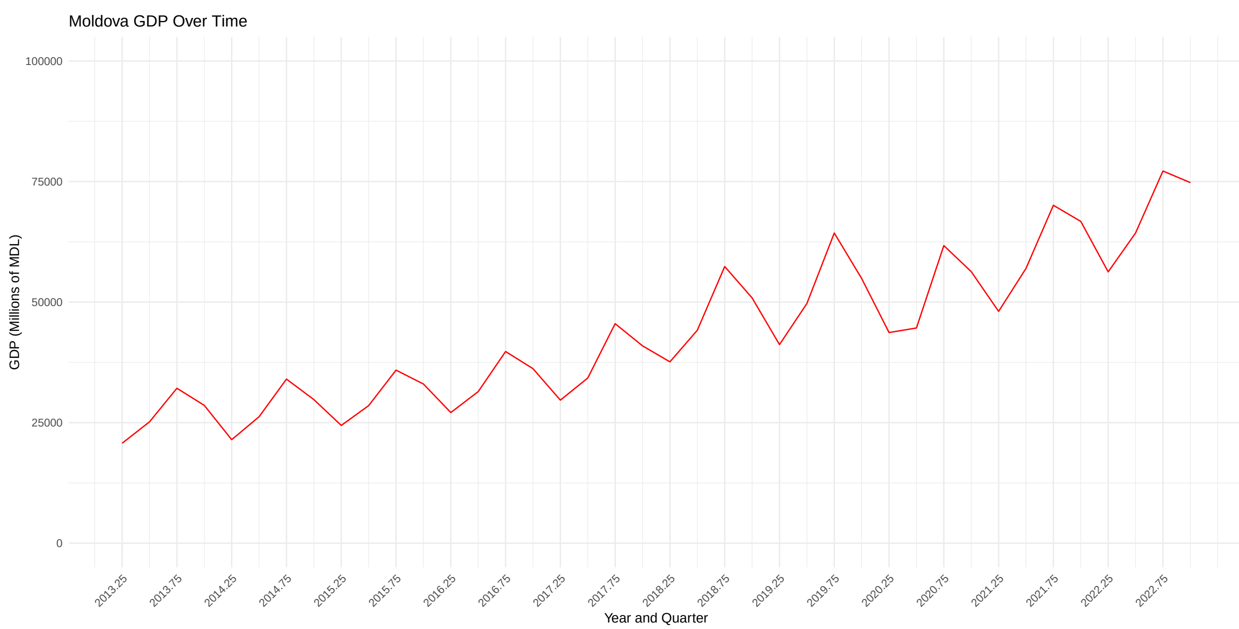


Figure 1: Moldova GDP Over Time

```
##
## Call:
## lm(formula = GDP_MDL ~ Date, data = gdp_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -10647  -4612   -402   3981  12840
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -9694152.5   713586.9  -13.59 3.65e-16 ***
## Date            4825.2      353.6   13.65 3.16e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6454 on 38 degrees of freedom
## Multiple R-squared:  0.8305, Adjusted R-squared:  0.8261
## F-statistic: 186.2 on 1 and 38 DF,  p-value: 3.165e-16
```

Now we will plot the relative change in quarterly GDP, expressed in percentage.

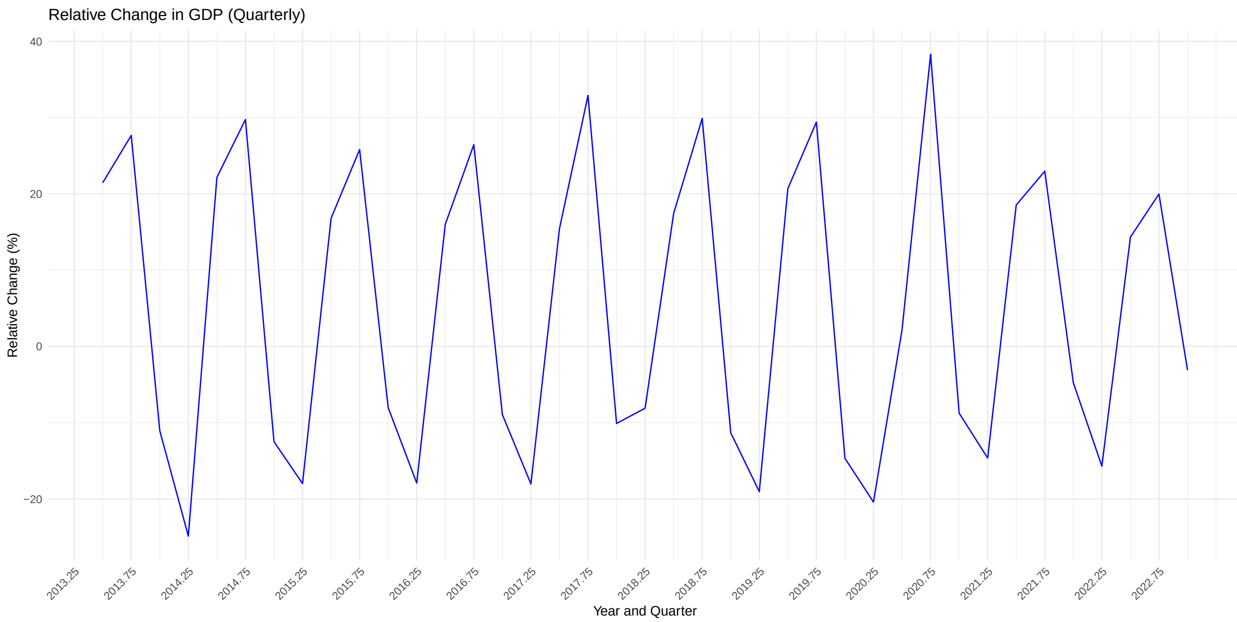


Figure 2: % Quarterly Change in GDP

Now we will plot the Year on Year (YoY) relative change in GDP.

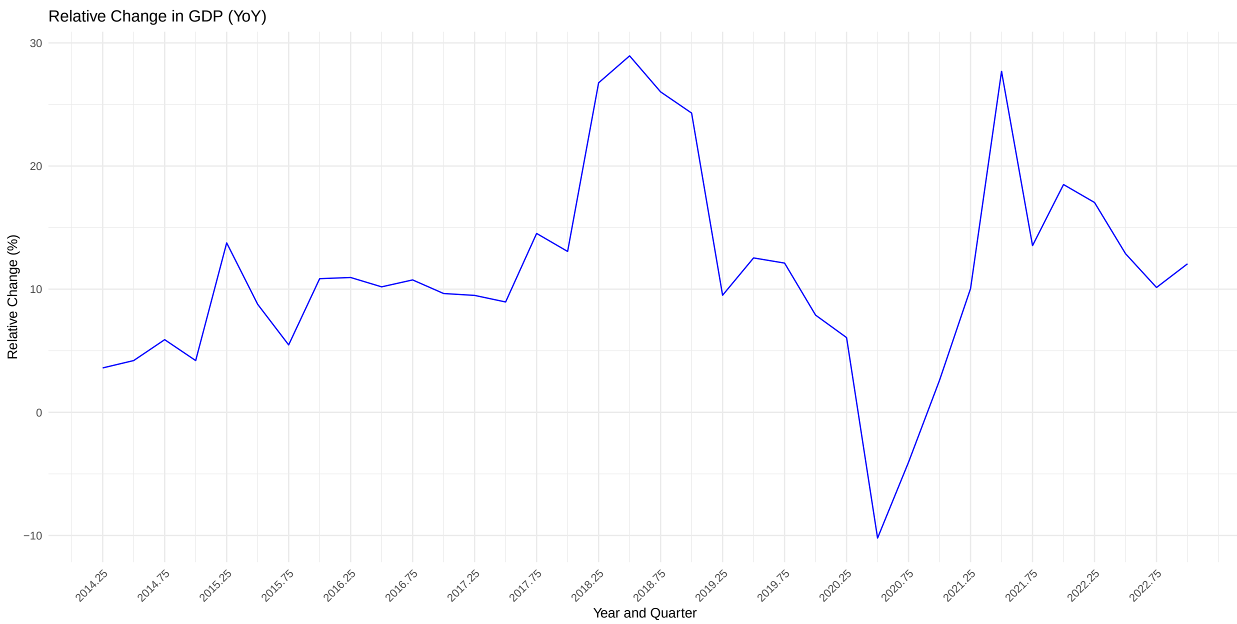


Figure 3: % YoY Change in GDP

The mean value of the YoY relative change in GDP is 11.3546117.

Now we will do the same process and data analysis but with the log of the GDP values, as well as a least squares fit.

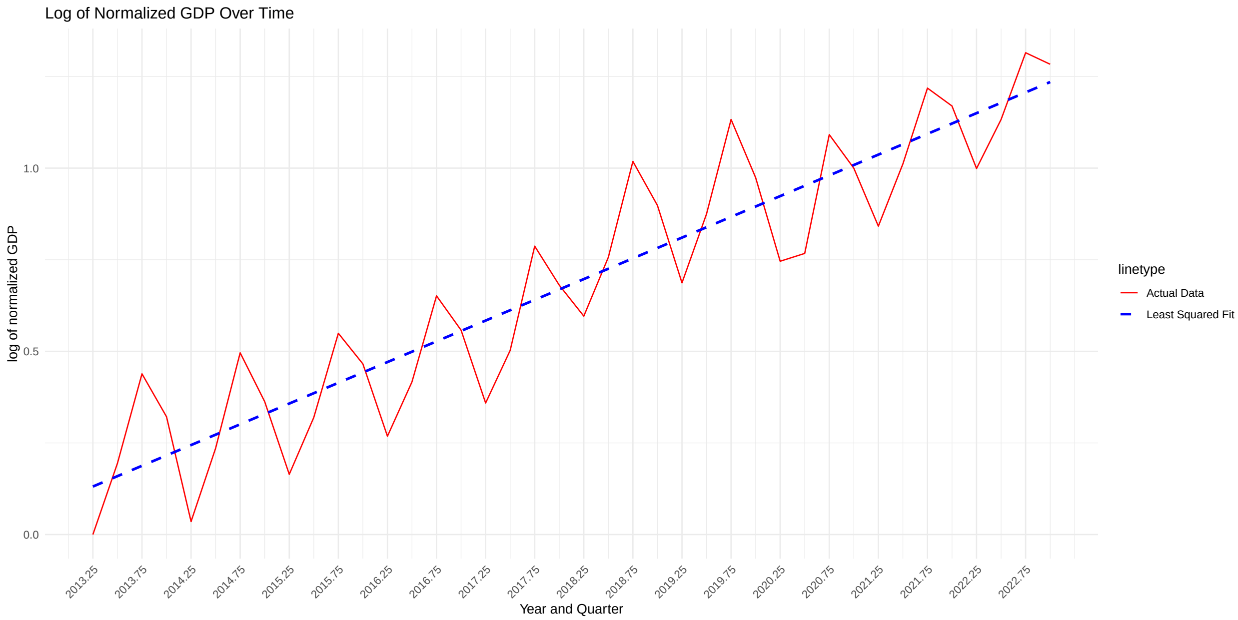


Figure 4: Log Moldova GDP Over Time with a LM

Below you can find the Quarterly Change in Log GDP.

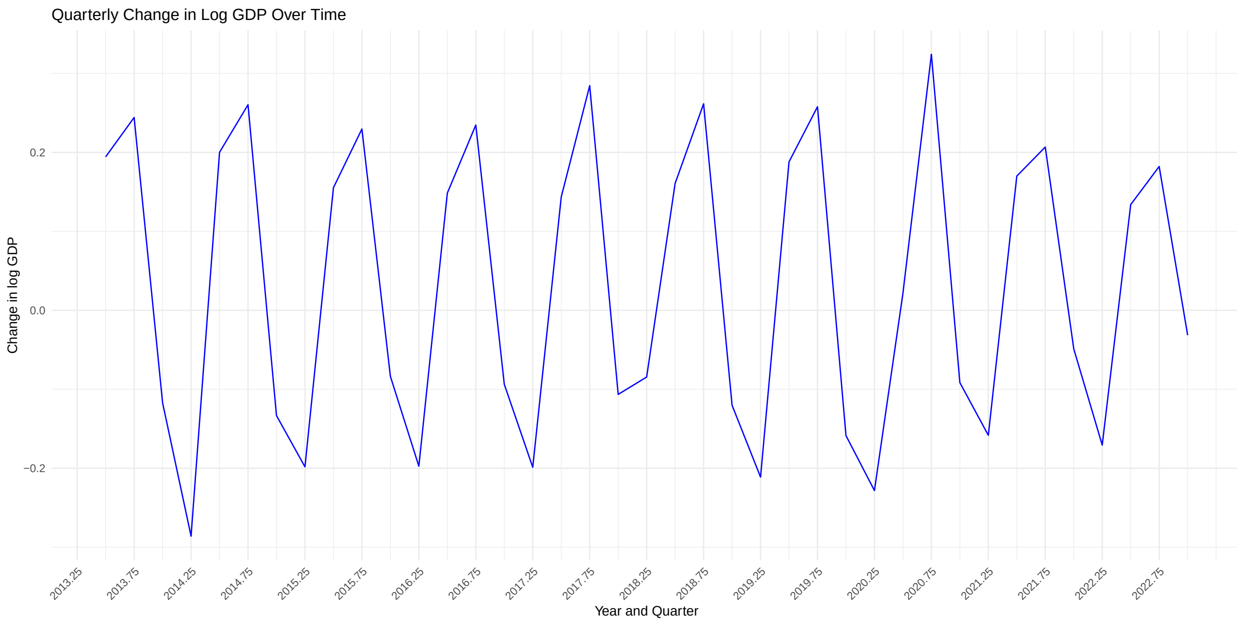


Figure 5: Quarterly Change in Log GDP

And the respective YoY Change in Log GDP

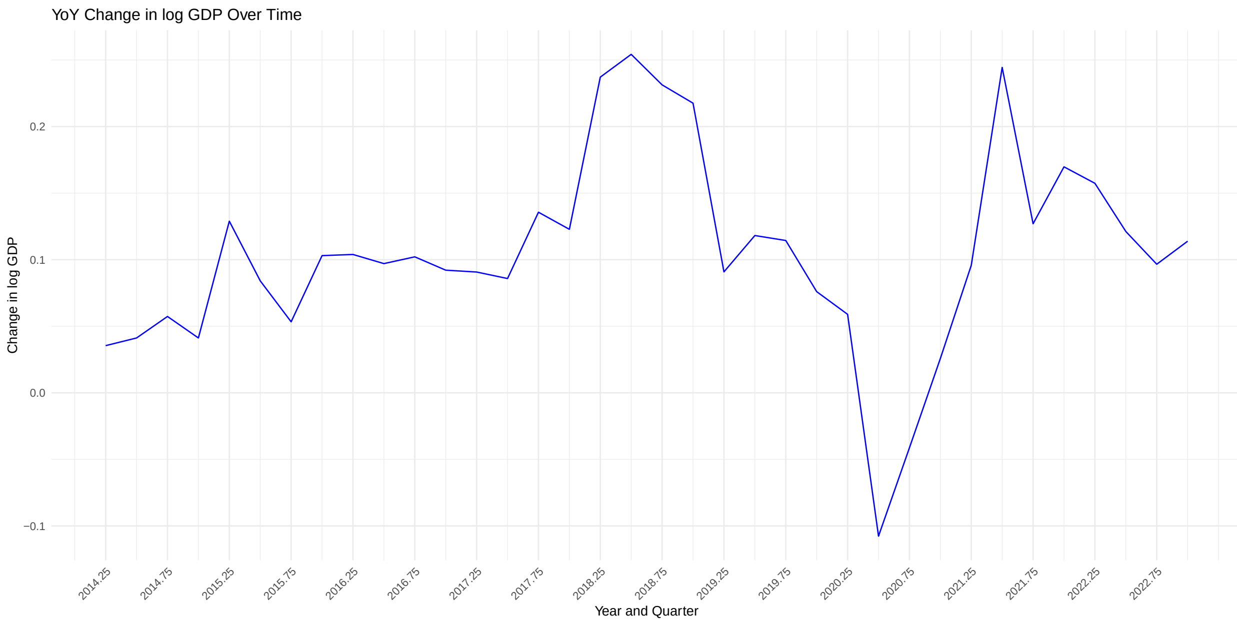


Figure 6: YoY Change in Log GDP

Now we will check the stationarity of the data.

```
##
##  Augmented Dickey-Fuller Test
##
## data:  residuals_diff_clean
## Dickey-Fuller = -3.1544, Lag order = 3, p-value = 0.124
## alternative hypothesis: stationary
```

We see that the data is not quite stationary.

Visualization of Google Trends data

The Google Trends data are divided in 20 main categories (with possible subcategories), out of which 10 were selected. The raw data represents the monthly relative value of the Google Trends within each category. That is, selecting the category, the region, and the time range (01.01.2013 - 31.12.2022 in our case) we obtain a number from 0 to 100 for every month.

To assure compatibility between the datasets, the Google Trends dataset was averaged by quarter to match the number of observations of GDP. Further in the research we could implement Mixed Frequence models that would use the maximum available data points.

That is, after processing the Google Trends dataset, for category i and time point t, we will represent the Google Trends value as $\mathrm{GT}_{i,t}$.

Below you will find an aggregate graph with the Google Trends categories plotted over the same period of time.

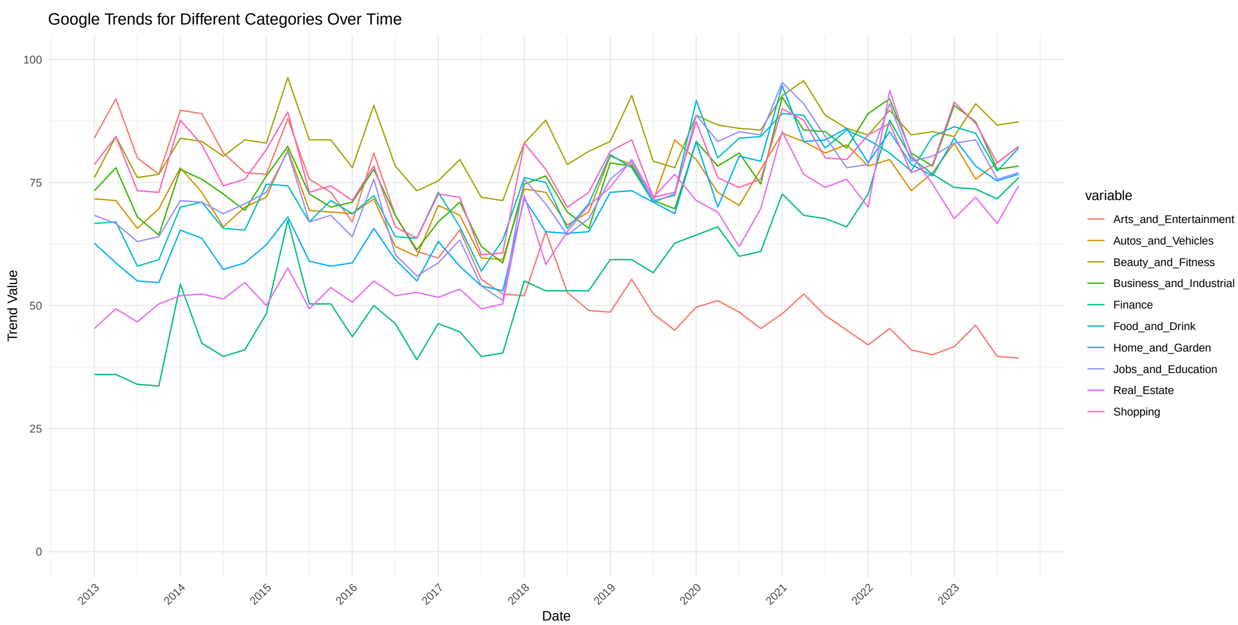


Figure 7: All Google Trends Categories Over Time

For further clarity, we will plot every graph separately.

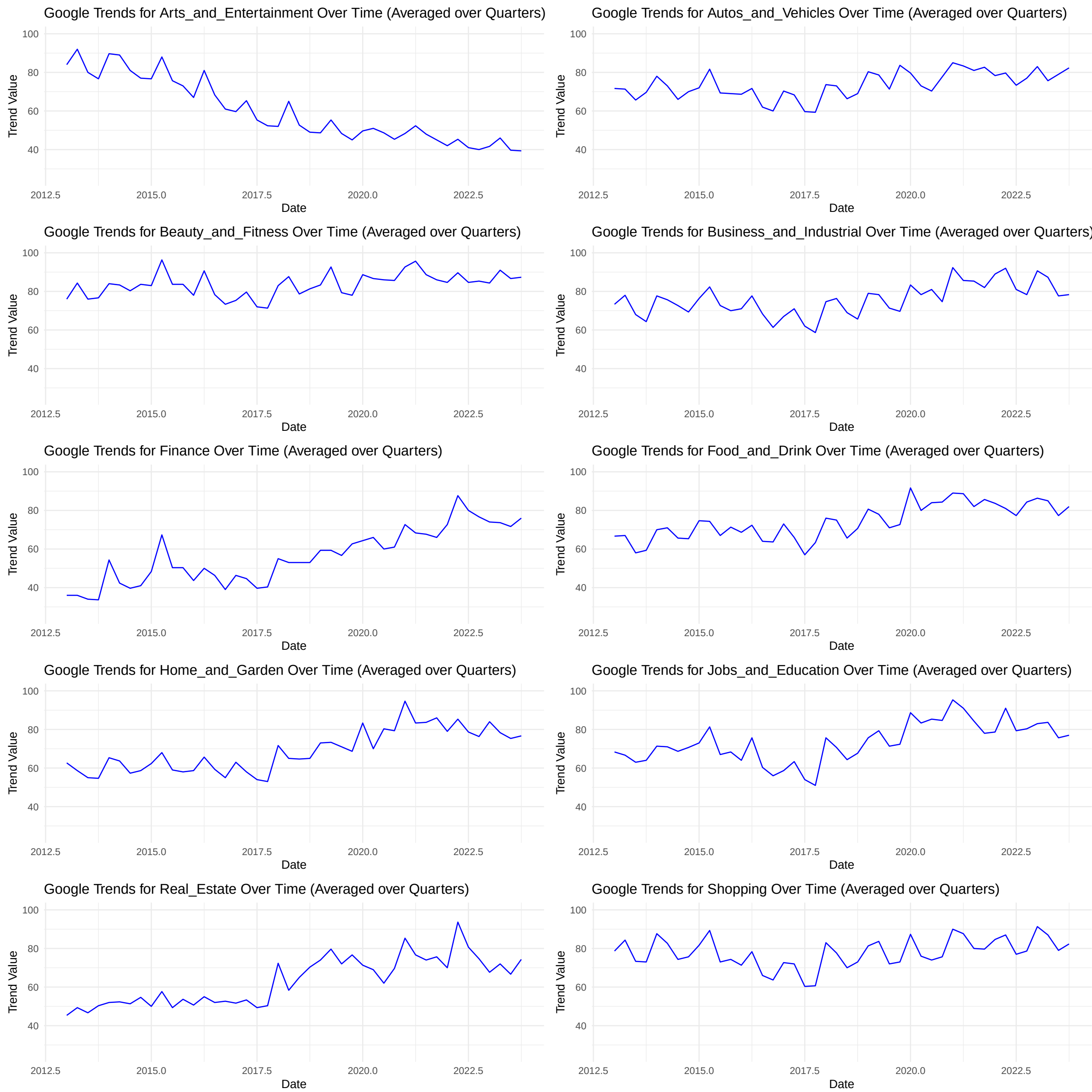


Figure 8: GT Categories Over Time

To further analyze the behavior of Google Trends data, we will visualize the relative difference in the Google Trends dataset (quarterly and yearly).

Below are the graphs of relative change in Google Trends over the quarters.

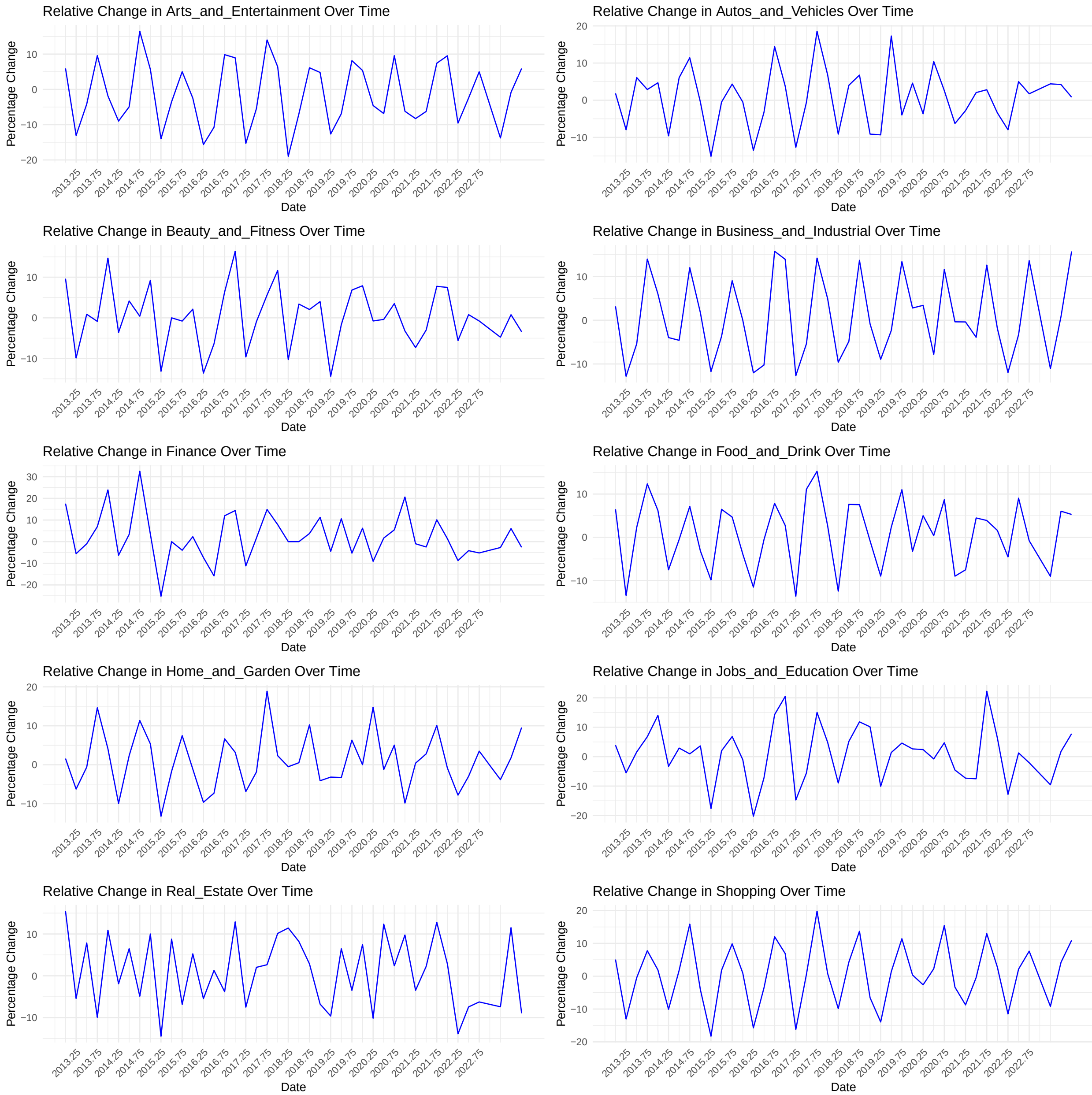


Figure 9: Quarterly Relative Chenge in GT

Now we will do the Year on Year relative change in Google Trends.

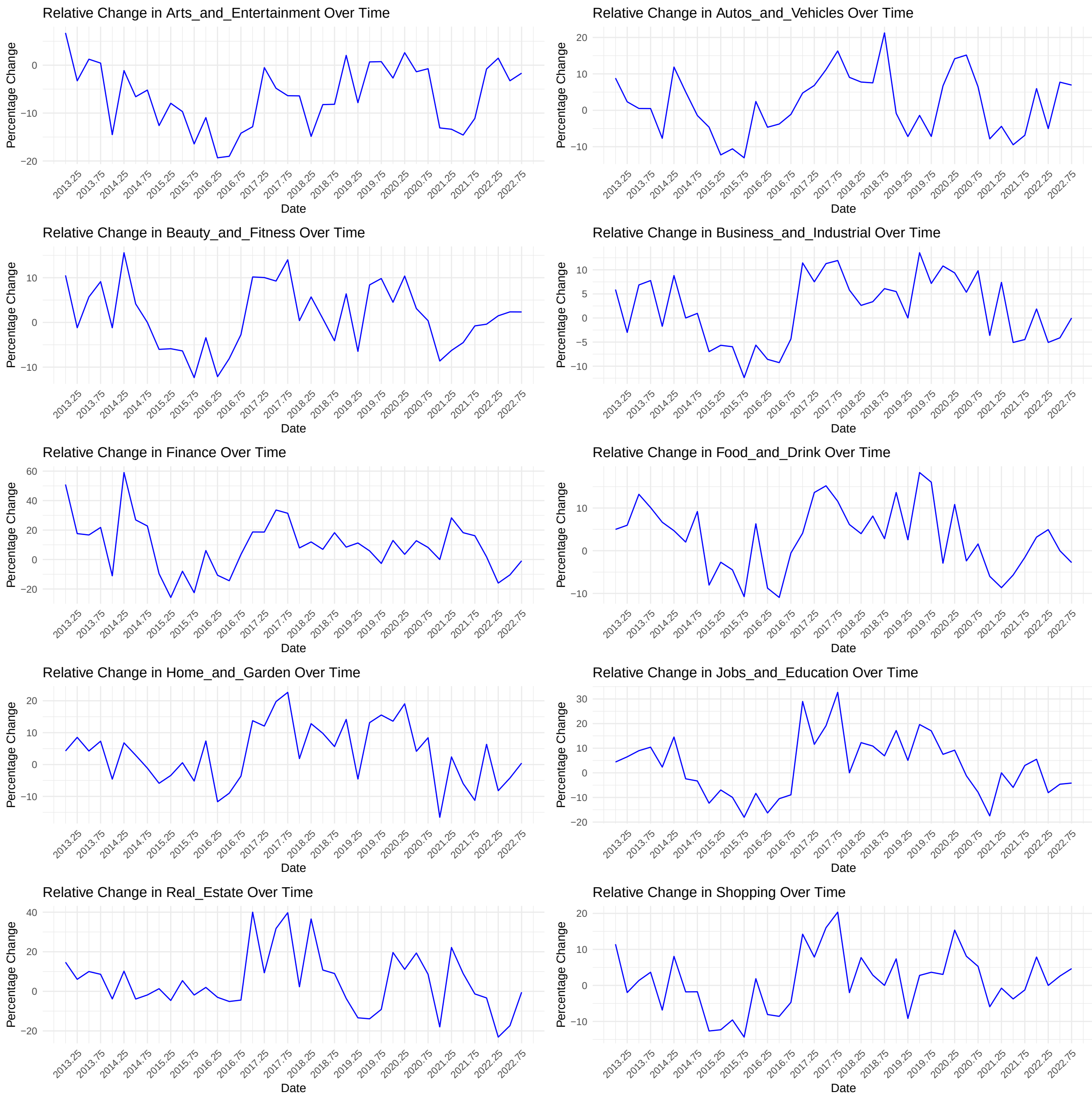


Figure 10: YoY % Change in GT

Now we will fit a least squares line to the GT graphs and plot the residuals.

The green dashed line represents the mean value of the residuals, which is close enough to zero.

Now we will plot the log of the GT data, as well as the least squares fit and the residuals, together with the R squared value for every category.

Afterwords, we would like to fit a least squares linear model and plot it as well as the residuals after we removed the seasonal component from the data, that is, the lag = 4 difference in the data.

```
## The R squared value for the time linear model
##  for category Arts_and_Entertainment is 0.898134935801906
## The R squared value for the time linear model
##  for category Autos_and_Vehicles is 0.292244542908991
## The R squared value for the time linear model
##  for category Beauty_and_Fitness is 0.186319077392161
## The R squared value for the time linear model
##  for category Business_and_Industrial is 0.298593836992014
## The R squared value for the time linear model
##  for category Finance is 0.763059670046091
## The R squared value for the time linear model
##  for category Food_and_Drink is 0.572292406075272
## The R squared value for the time linear model
##  for category Home_and_Garden is 0.64760468235368
## The R squared value for the time linear model
##  for category Jobs_and_Education is 0.346790220092301
## The R squared value for the time linear model
##  for category Real_Estate is 0.714058949228139
## The R squared value for the time linear model
##  for category Shopping is 0.0781262013412593
```

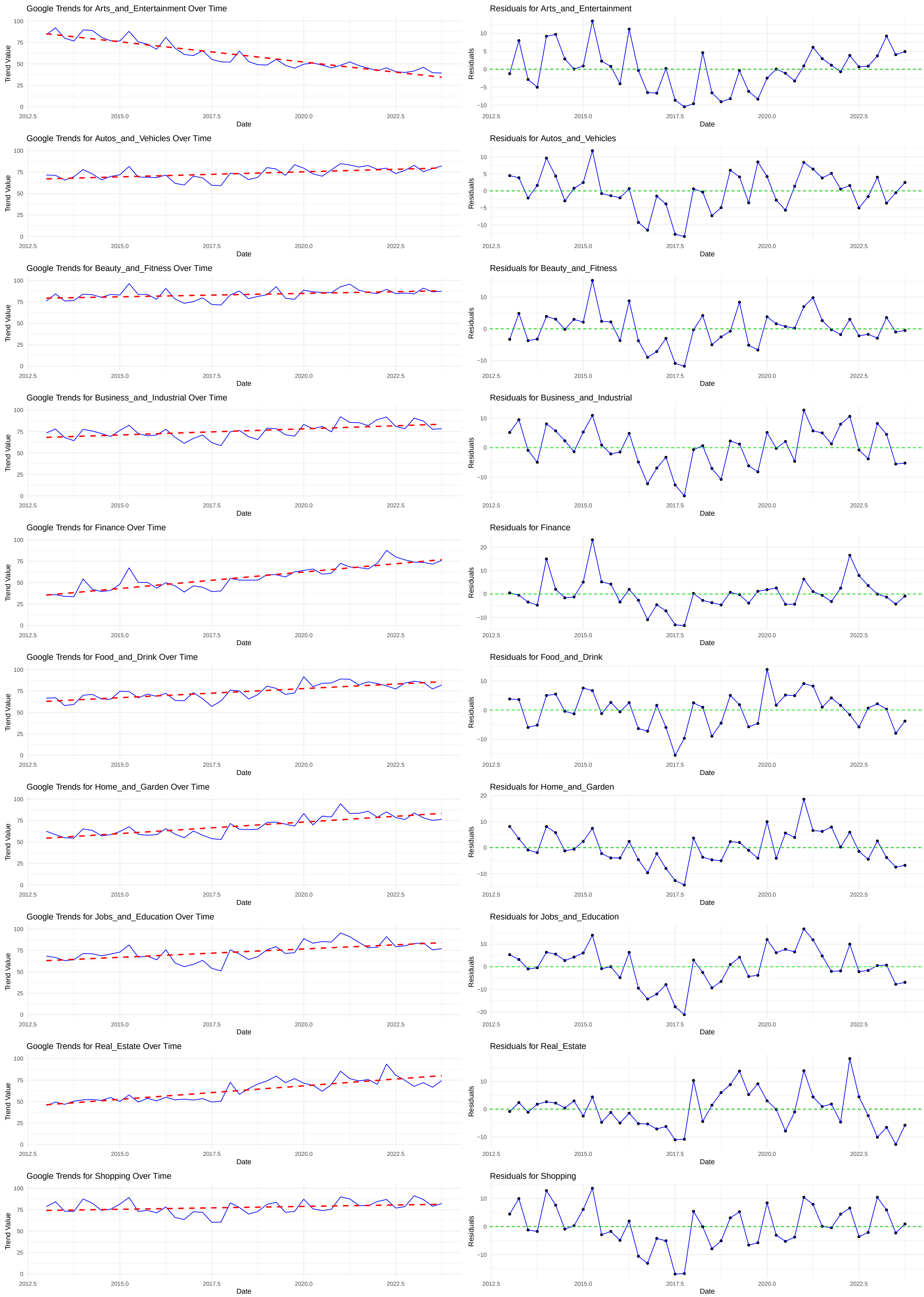


Figure 11: Linear Fit and Residuals

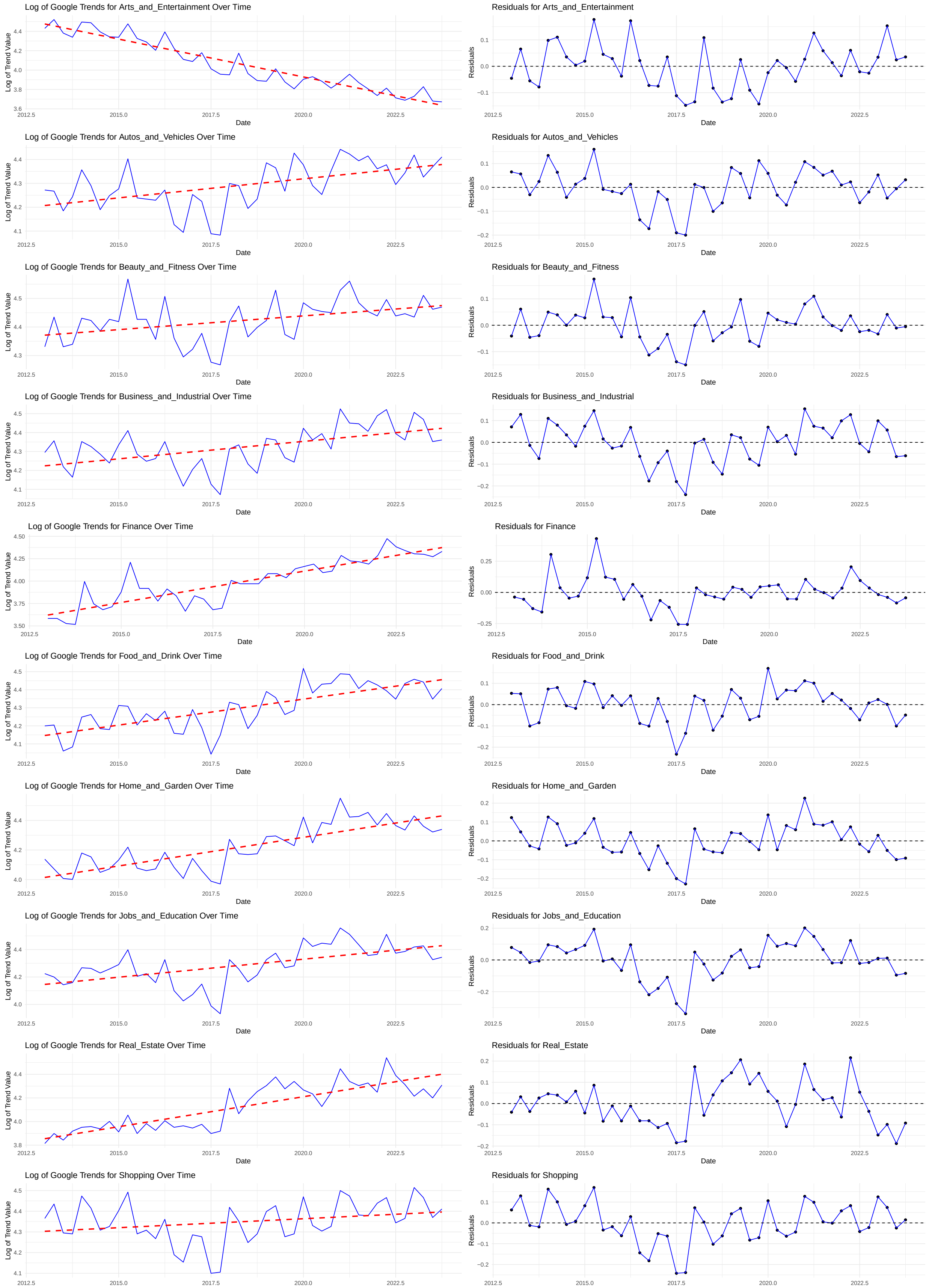


Figure 12: Linear Fit and Residuals of Log GT

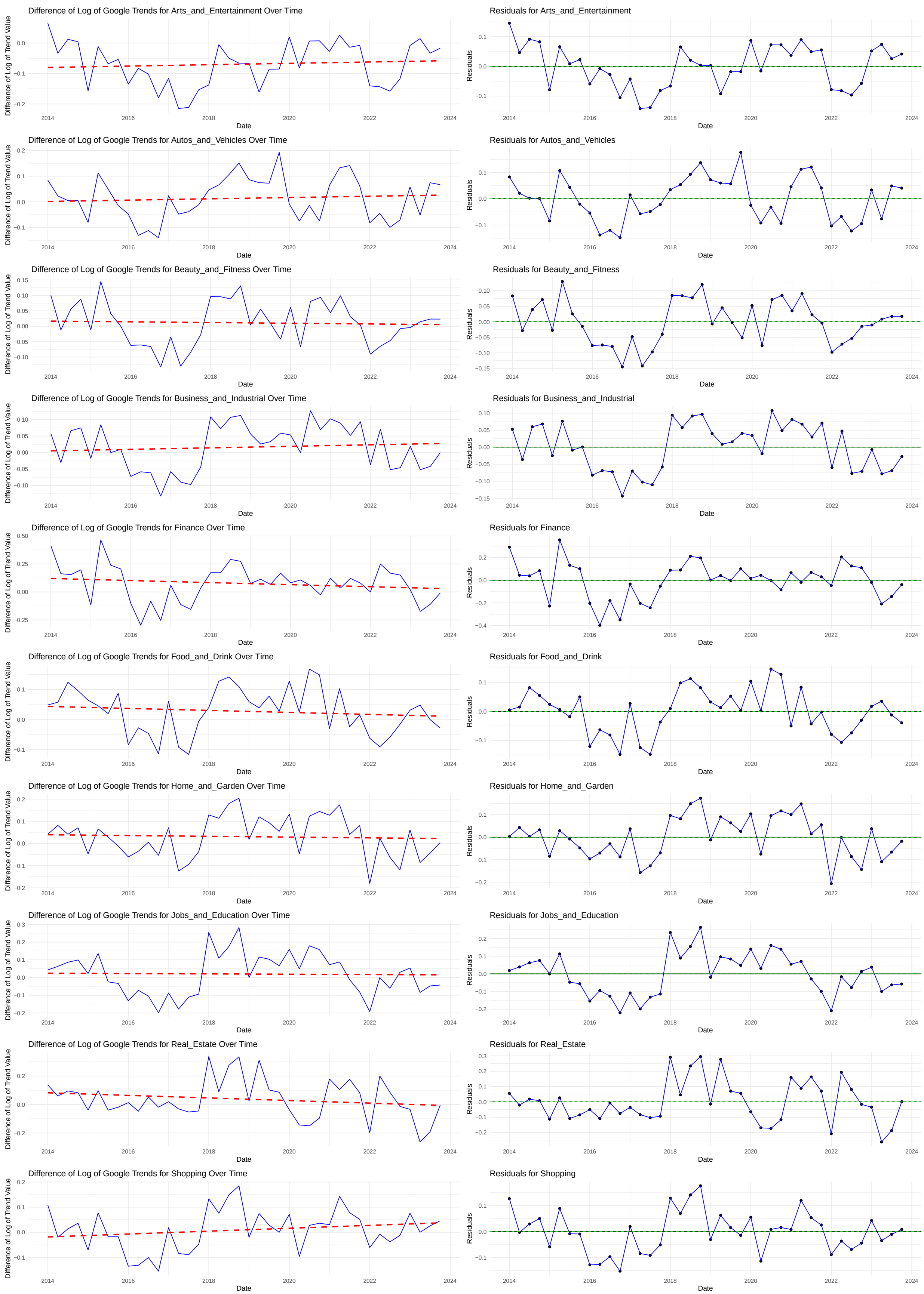


Figure 13: Linear Fit and Residuals for Log GT

Overlapping

Below you can find the overlap between the graphs of normalized values of GDP and GT

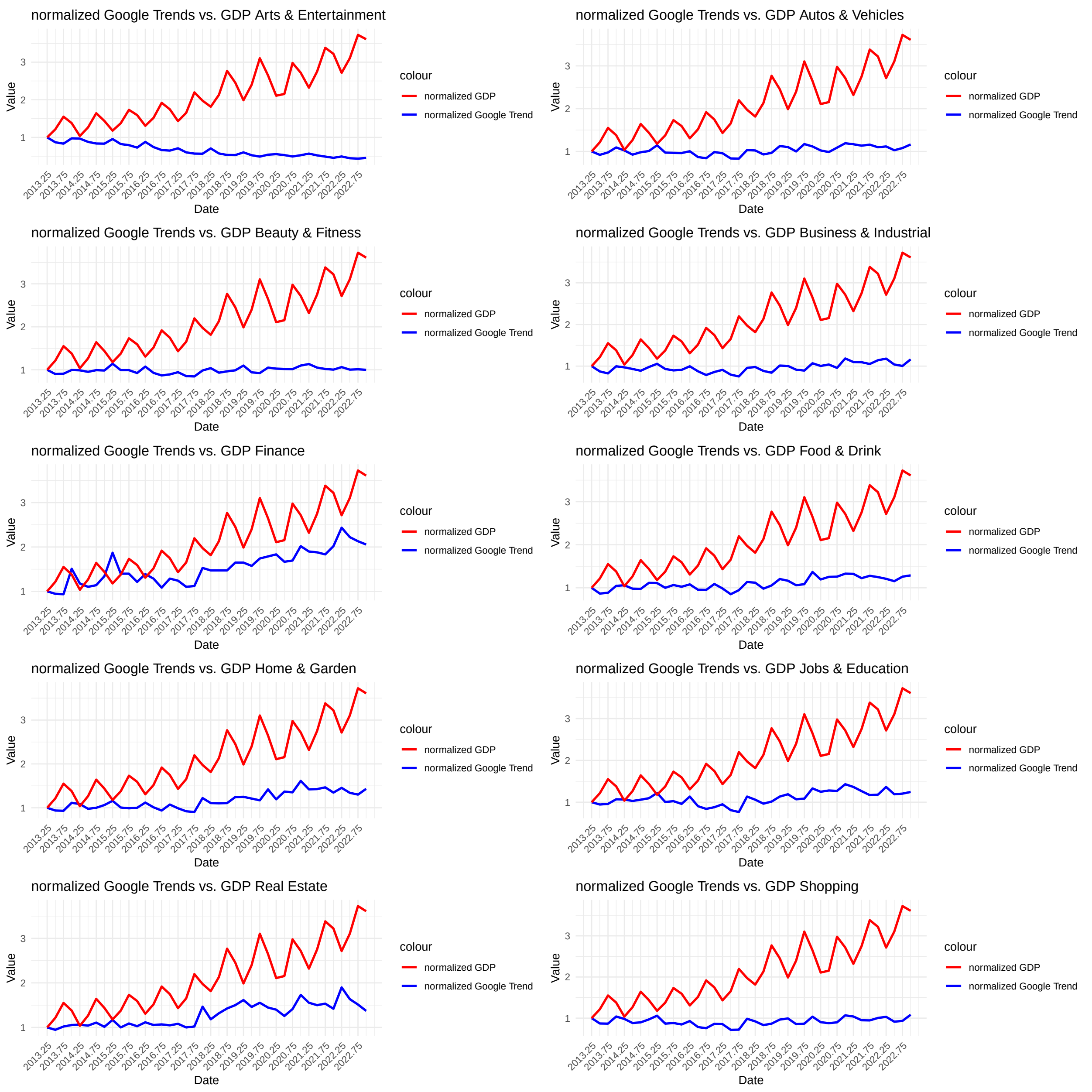


Figure 14: Normalized GDP vs. GT

Below you can find the overlap between the graphs of quarterly relative change of Google Trends and quarterly relative change of GDP.

Now we will plot the overlap of the Log of the normalized values of GT and GDP.

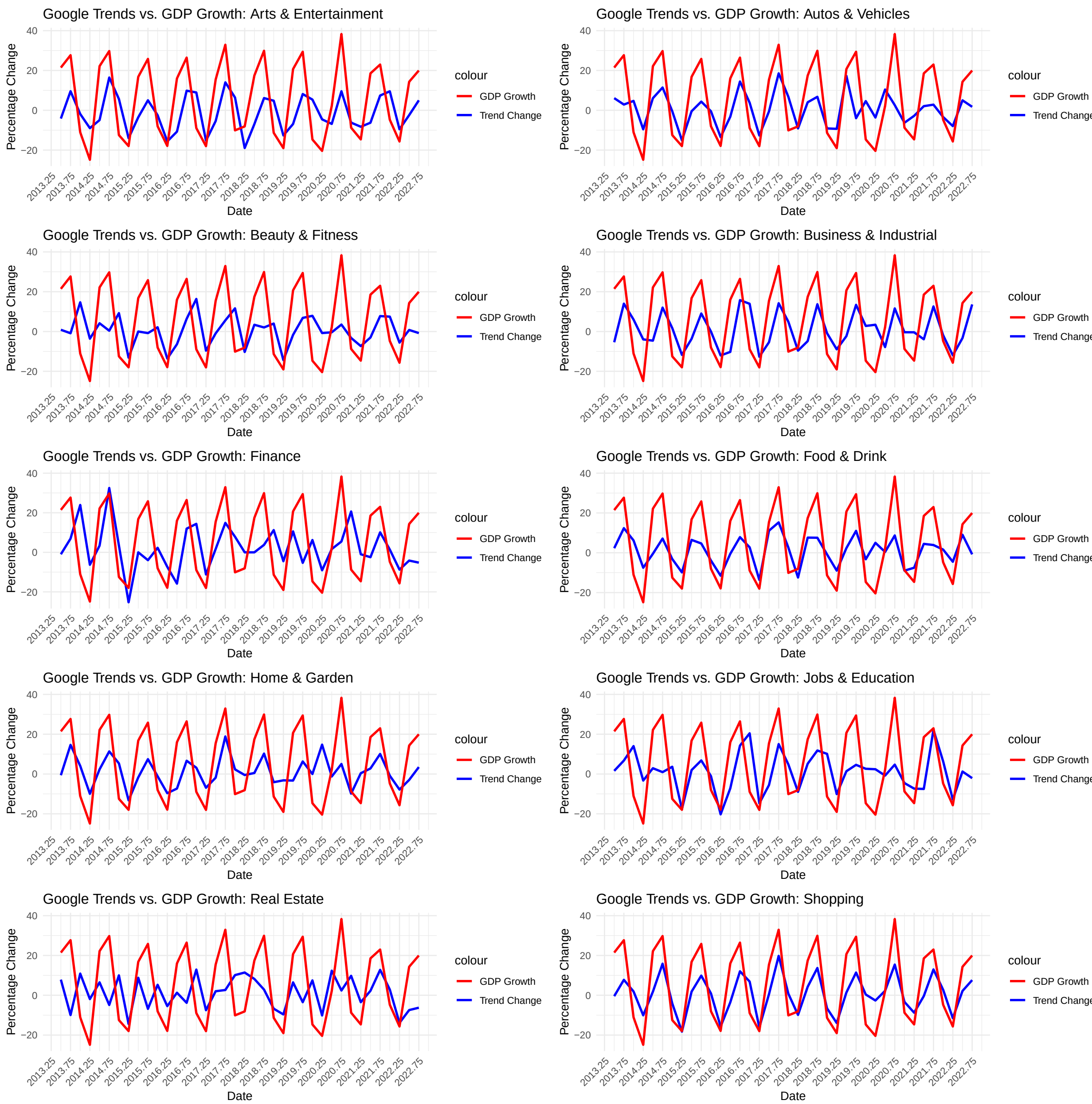


Figure 15: Quarterly % Change GDP vs. GT

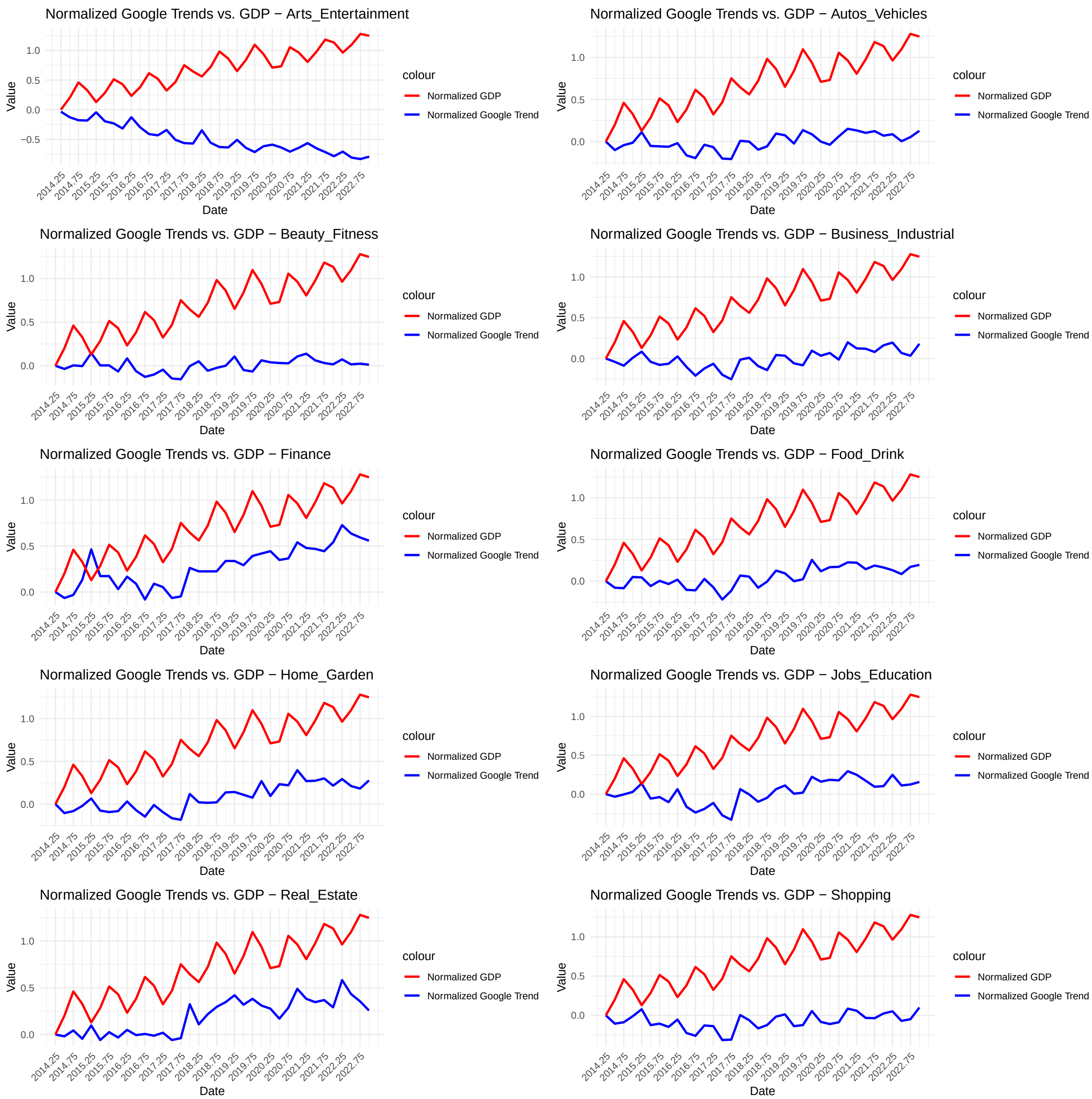


Figure 16: Log of Normalized GDP vs. GT

Scatter Plots

Now we will plot the quarterly relative change in Google Trends versus the quarterly relative change in GDP.

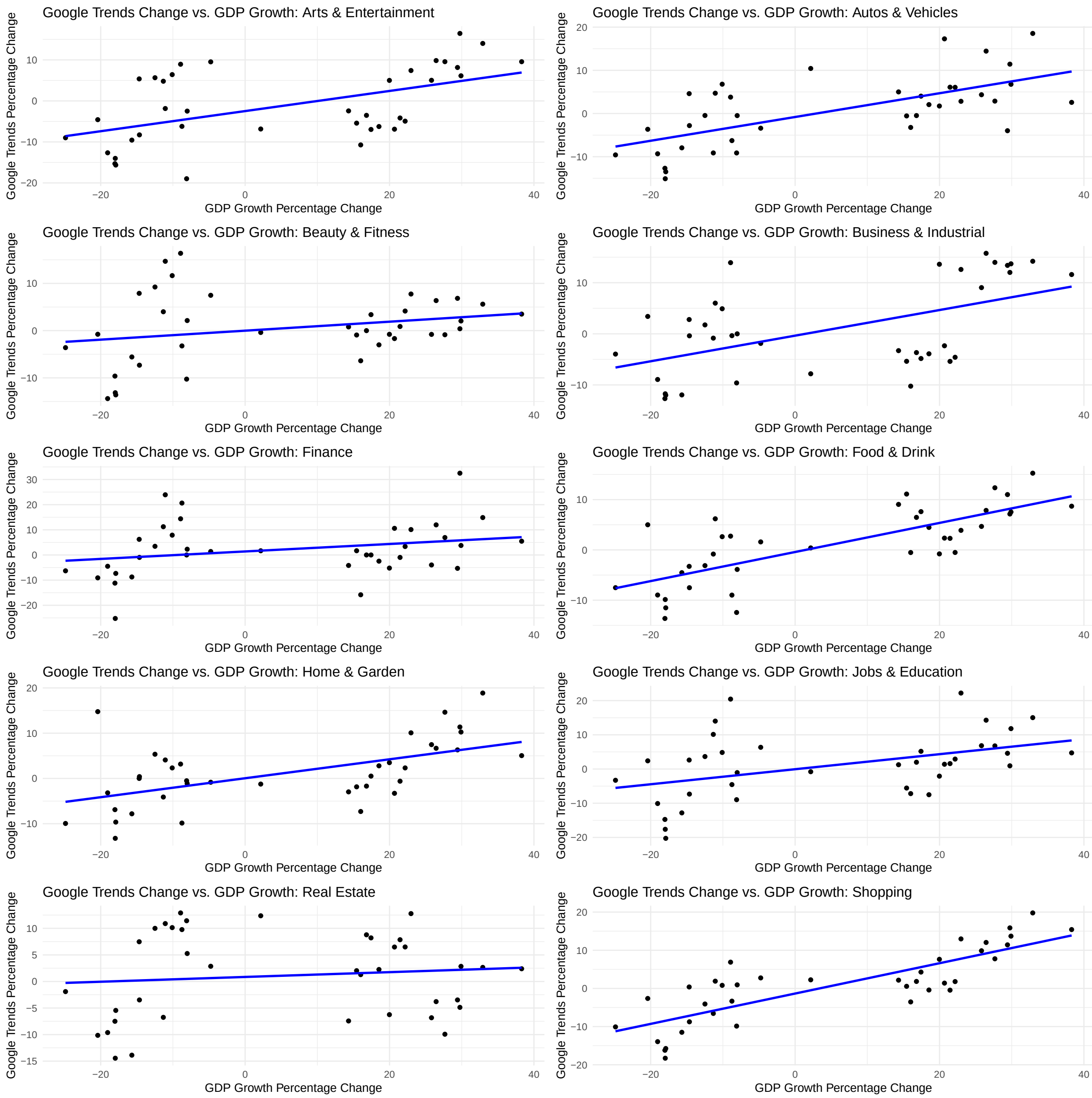


Figure 17: Quarterly % Change in GT versus % Change in GDP

We identify that the points in every graph are divided in two clusters. Thus, it is rationale to analyze and do a linear fit for every cluster separately.

Further, let's plot the YoY relative change in GT versus the YoY relative change in GDP.

Now, to establish a relationship between these values, we will plot the Log-Log graph of the normalized values of Google Trends and GDP.

```
## [1] "The R squared value for category Arts_Entertainment is 0.911220075585741"
```

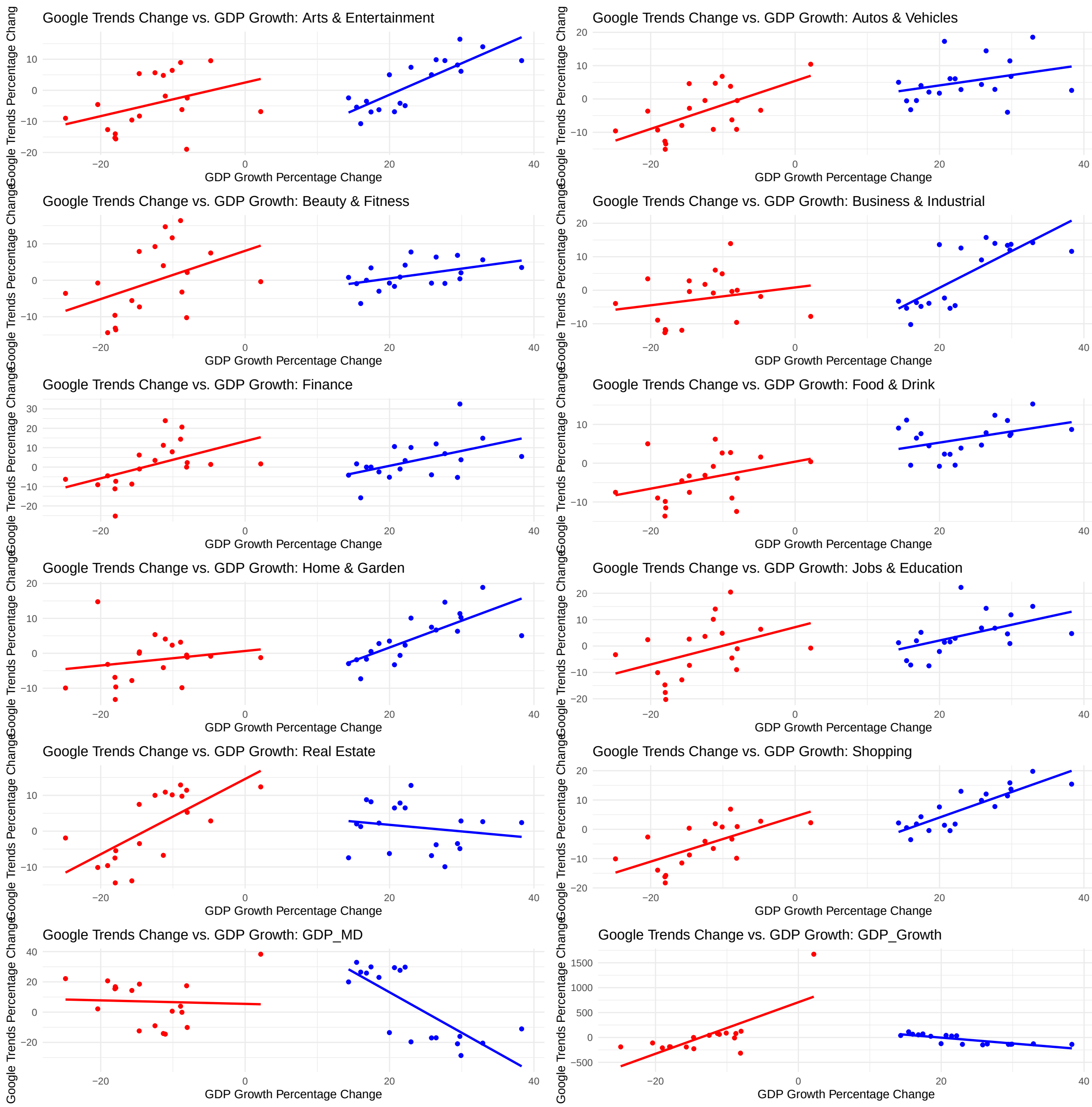


Figure 18: Divided by clusters

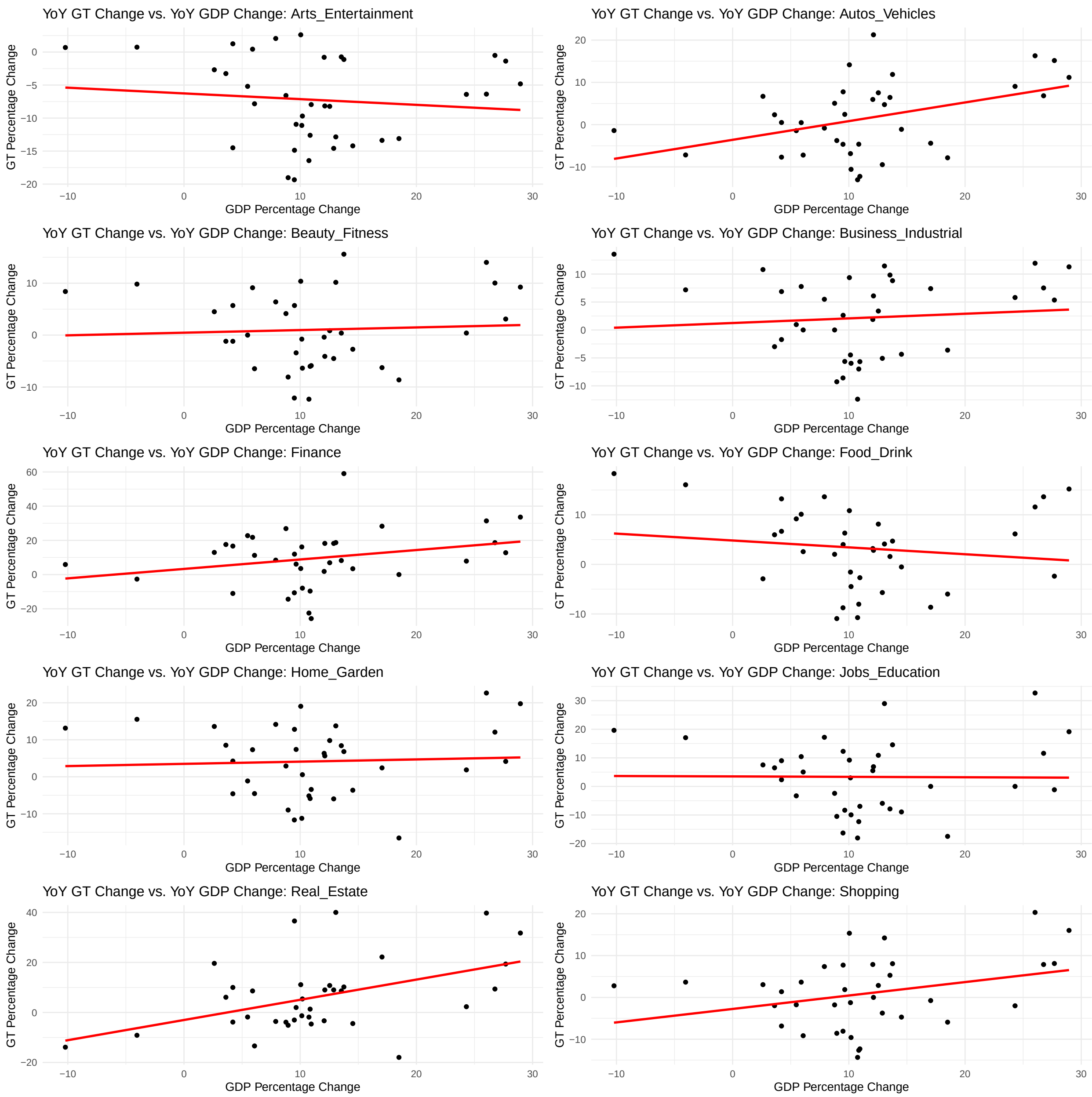


Figure 19: YoY % Change in GT vs. YoY % Change in GDP

```
## [1] "The R squared value for category Autos_Vehicles is 0.220672996345828"
## [1] "The R squared value for category Beauty_Fitness is 0.0264465997520018"
## [1] "The R squared value for category Business_Industrial is 0.130234505741318"
## [1] "The R squared value for category Finance is 0.547001032621618"
## [1] "The R squared value for category Food_Drink is 0.426611598453415"
## [1] "The R squared value for category Home_Garden is 0.509093984356865"
## [1] "The R squared value for category Jobs_Education is 0.216358583108777"
## [1] "The R squared value for category Real_Estate is 0.643866940708585"
## [1] "The R squared value for category Shopping is 0.0216254833557737"
```

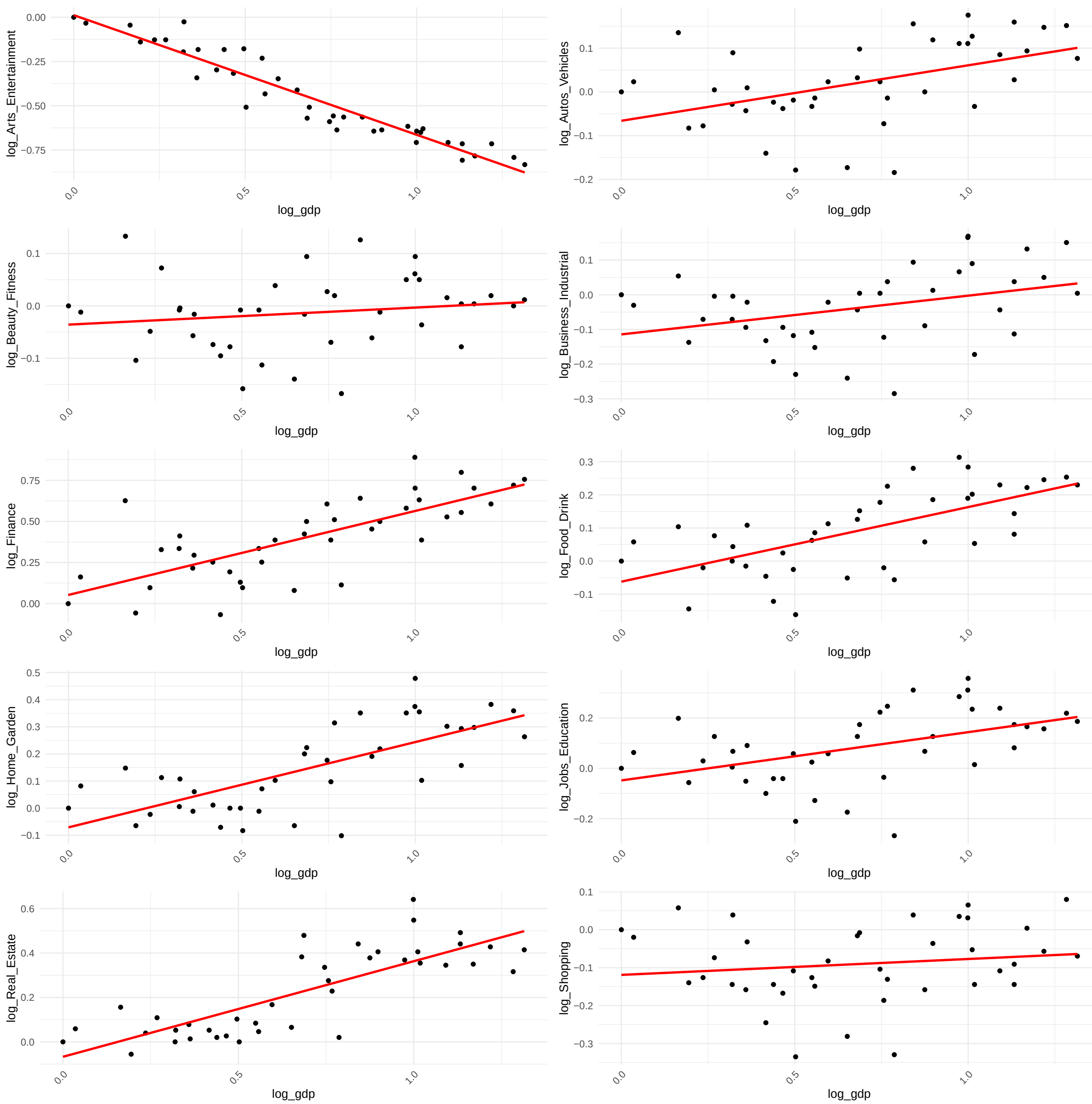


Figure 20: Log-Log Graph of GT vs. GDP

Data Modelling

Now, we performed a Principal Component Analysis (PCA) to determine which categories of search contribute most to the variance, if we interpret the data as samples from a multivariate distribution.

```
## Importance of components:
##                           PC1    PC2     PC3     PC4     PC5     PC6     PC7
## Standard deviation     2.7335 1.2414 0.51922 0.48862 0.42944 0.33958 0.26934
## Proportion of Variance 0.7472 0.1541 0.02696 0.02387 0.01844 0.01153 0.00725
## Cumulative Proportion  0.7472 0.9013 0.92825 0.95213 0.97057 0.98210 0.98935
##                            PC8     PC9    PC10
## Standard deviation     0.22853 0.18049 0.14715
## Proportion of Variance 0.00522 0.00326 0.00217
## Cumulative Proportion  0.99458 0.99783 1.00000
```

```
##      PC1                PC2                    PC3
## [1,] "Home & Garden"    "Arts & Entertainment" "Beauty & Fitness"
## [2,] "Jobs & Education" "Shopping"             "Autos & Vehicles"
## [3,] "Food & Drink"     "Beauty & Fitness"     "Shopping"
```

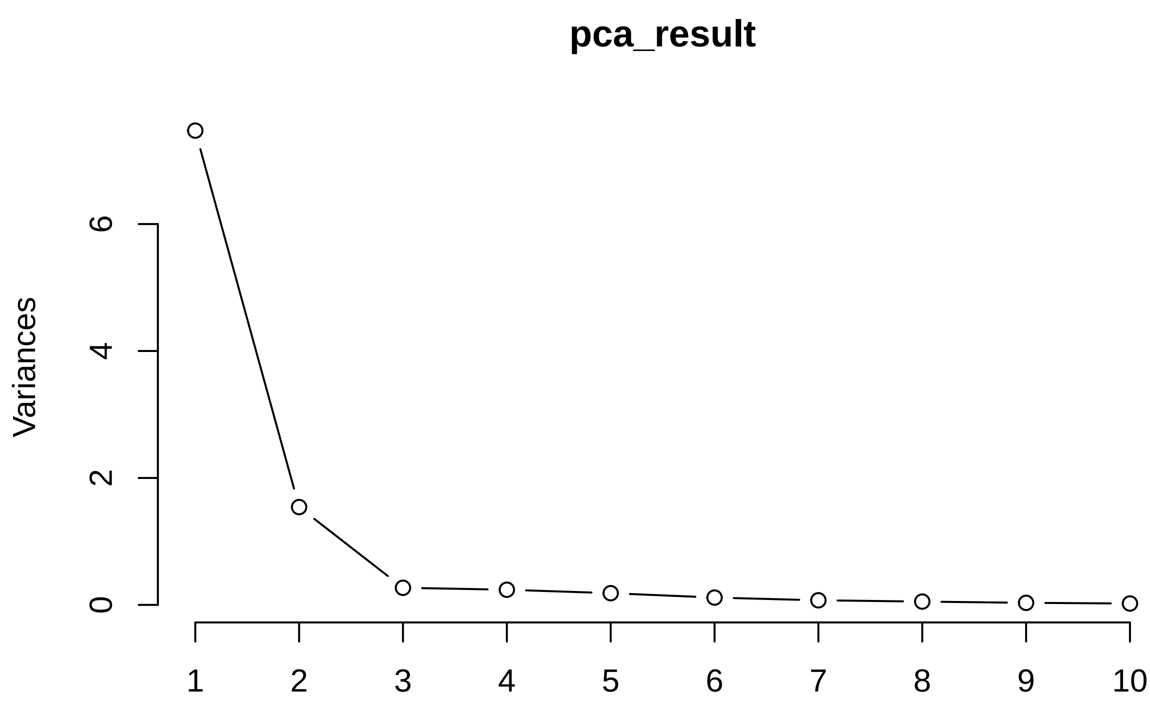


Figure 21: PCA Cummulative variance by component

```
##
## Call:
```

```
## lm(formula = GDP_MDL ~ PC1 + PC2 + PC3, data = pca_scores)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -10446  -4225   -159   2962  15100
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  43642.2      899.7  48.505  < 2e-16 ***
## PC1           3641.4      333.4  10.923 5.57e-13 ***
## PC2           8115.0      734.0  11.056 3.98e-13 ***
## PC3          -5776.9     1755.0  -3.292  0.00224 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5690 on 36 degrees of freedom
## Multiple R-squared:  0.8752, Adjusted R-squared:  0.8648
## F-statistic: 84.13 on 3 and 36 DF,  p-value: 2.486e-16
```

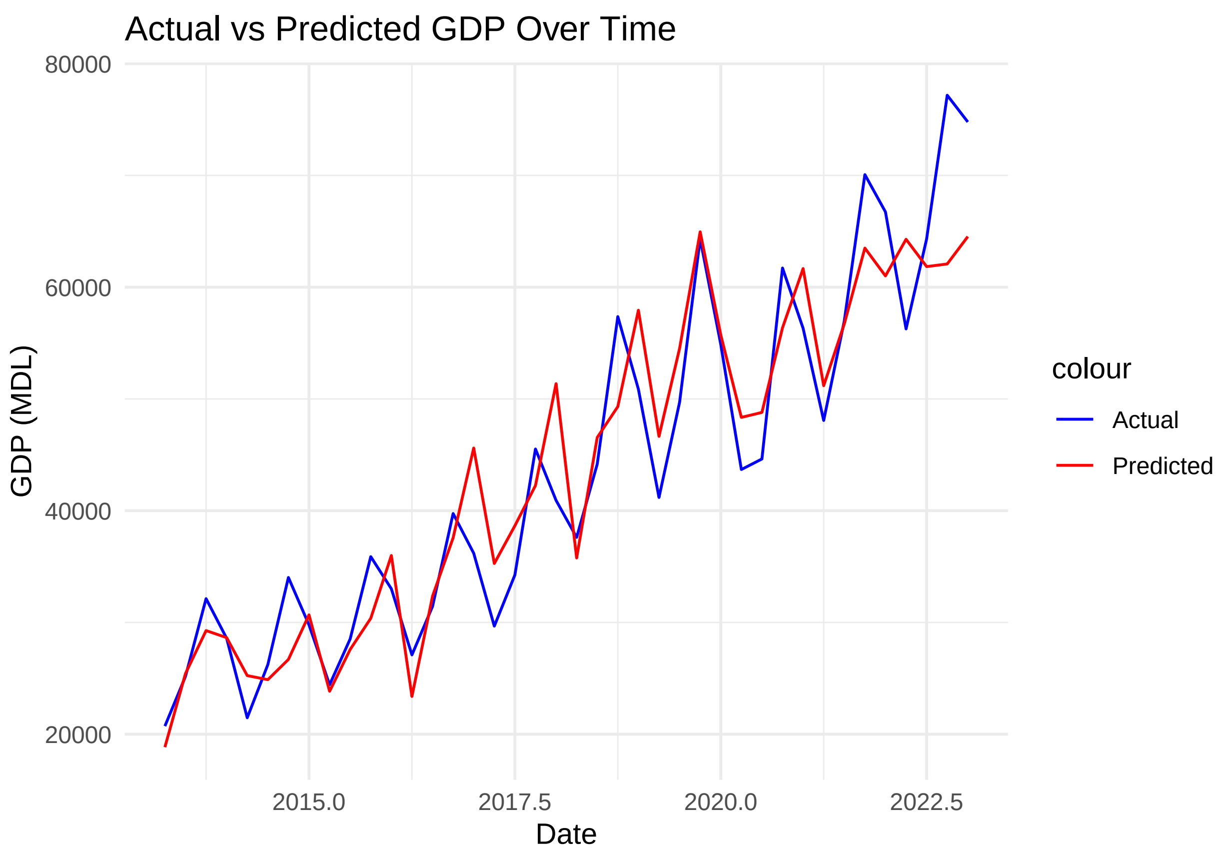


Figure 22: Figure 1: Moldova GDP Over Time

Now we will do a PLS regression.

```
## Data:    X dimension: 40 11
##  Y dimension: 40 1
```

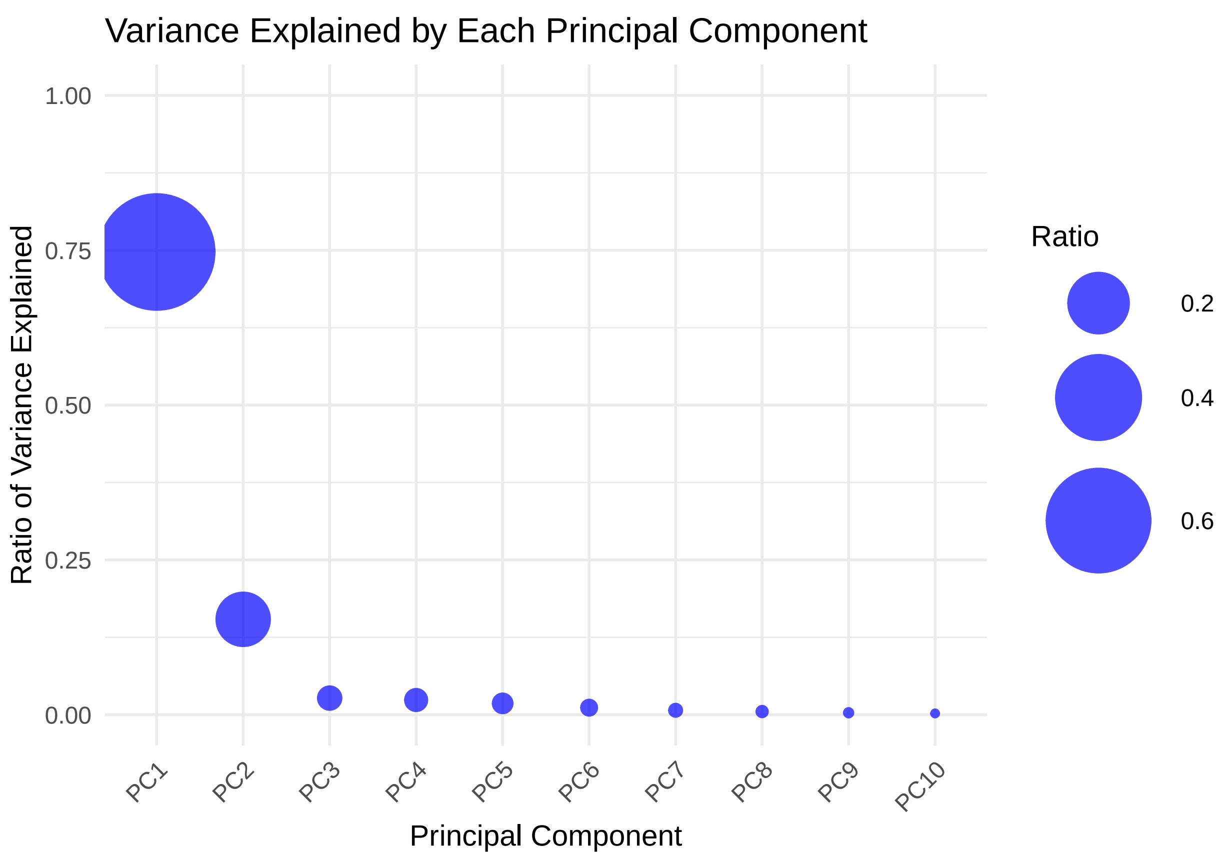


Figure 23: Figure 1: Moldova GDP Over Time

```
## Fit method: kernelpls
## Number of components considered: 3
##
## VALIDATION: RMSEP
## Cross-validated using 10 random segments.
##        (Intercept)  1 comps  2 comps  3 comps
## CV           1.013   0.5948   0.2764   0.1520
## adjCV        1.013   0.5928   0.2755   0.1511
##
## TRAINING: % variance explained
##                    1 comps  2 comps  3 comps
## X                    69.32    89.89    92.53
## standardized_gdp     69.37    93.41    98.07
```

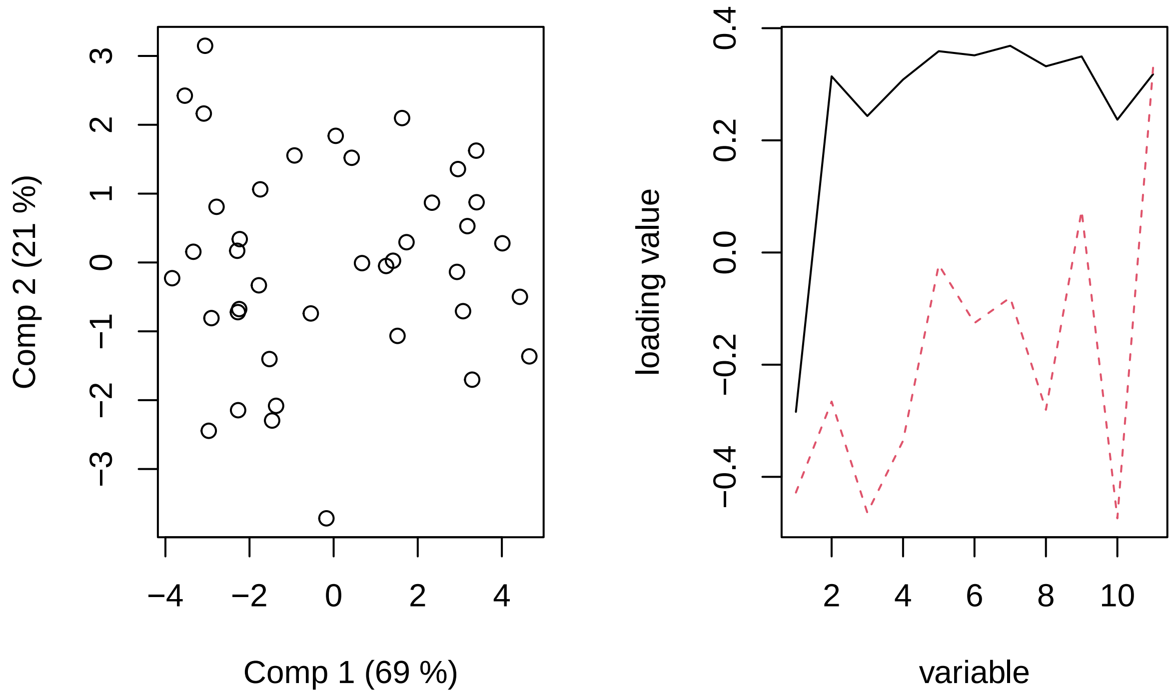


Figure 24: Figure 1: Moldova GDP Over Time

```
## [1] 0.9791137
```

Now we will model $\log GDP_T \sim \log GT_t$

```
## [1] 0.04990361
## [1] 6.582382e-05
## [1] 0.16326
## [1] 0.1177884
```

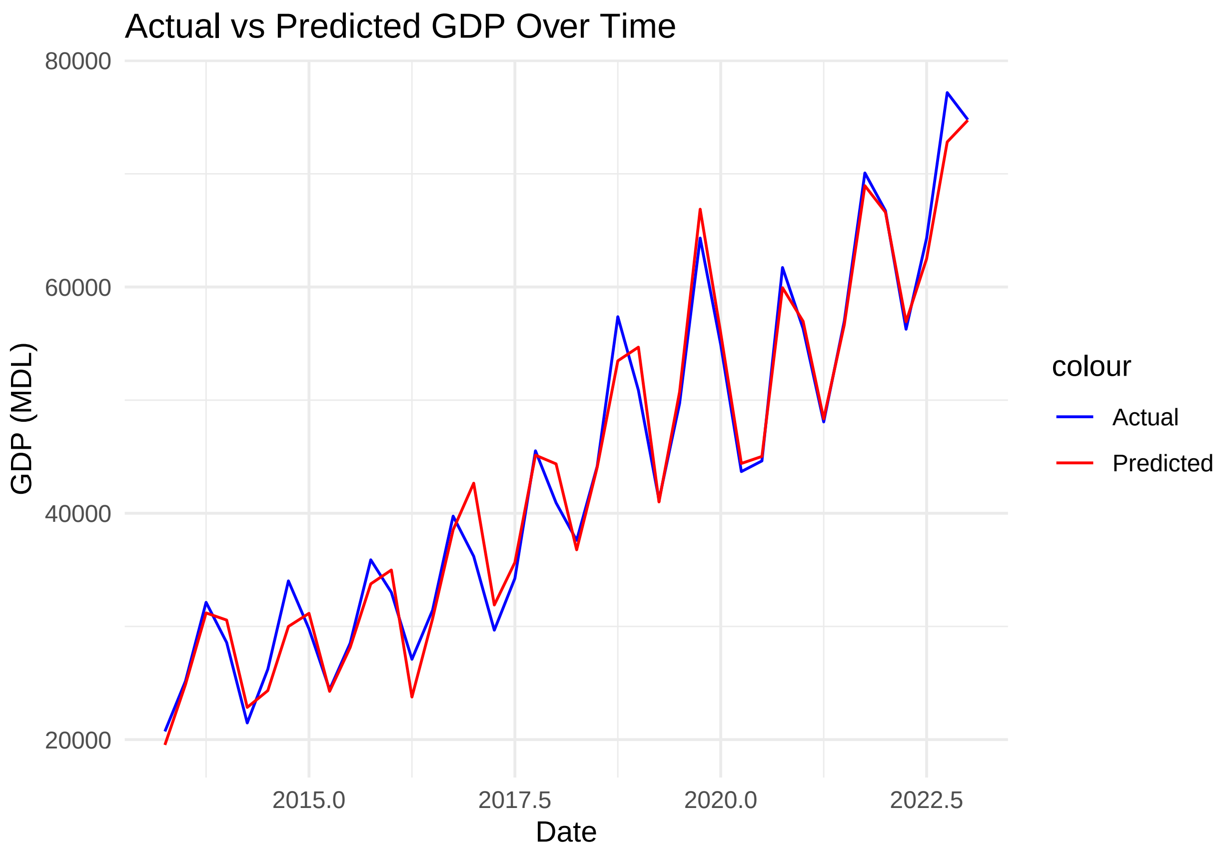


Figure 25: Figure 1: Moldova GDP Over Time

```
## [1] 0.01661224
## [1] 0.01076655
## [1] 0.01183329
## [1] 0.04642366
## [1] 0.001930843
## [1] 0.03177104
```

Now we will implement a regressive model, where we fit the linear model $\log GDP_t \sim \log GT_t + \log GT_{t-1}$ model after removing the trend and looking at residuals.

Now we will add the trend back to the predicted values.

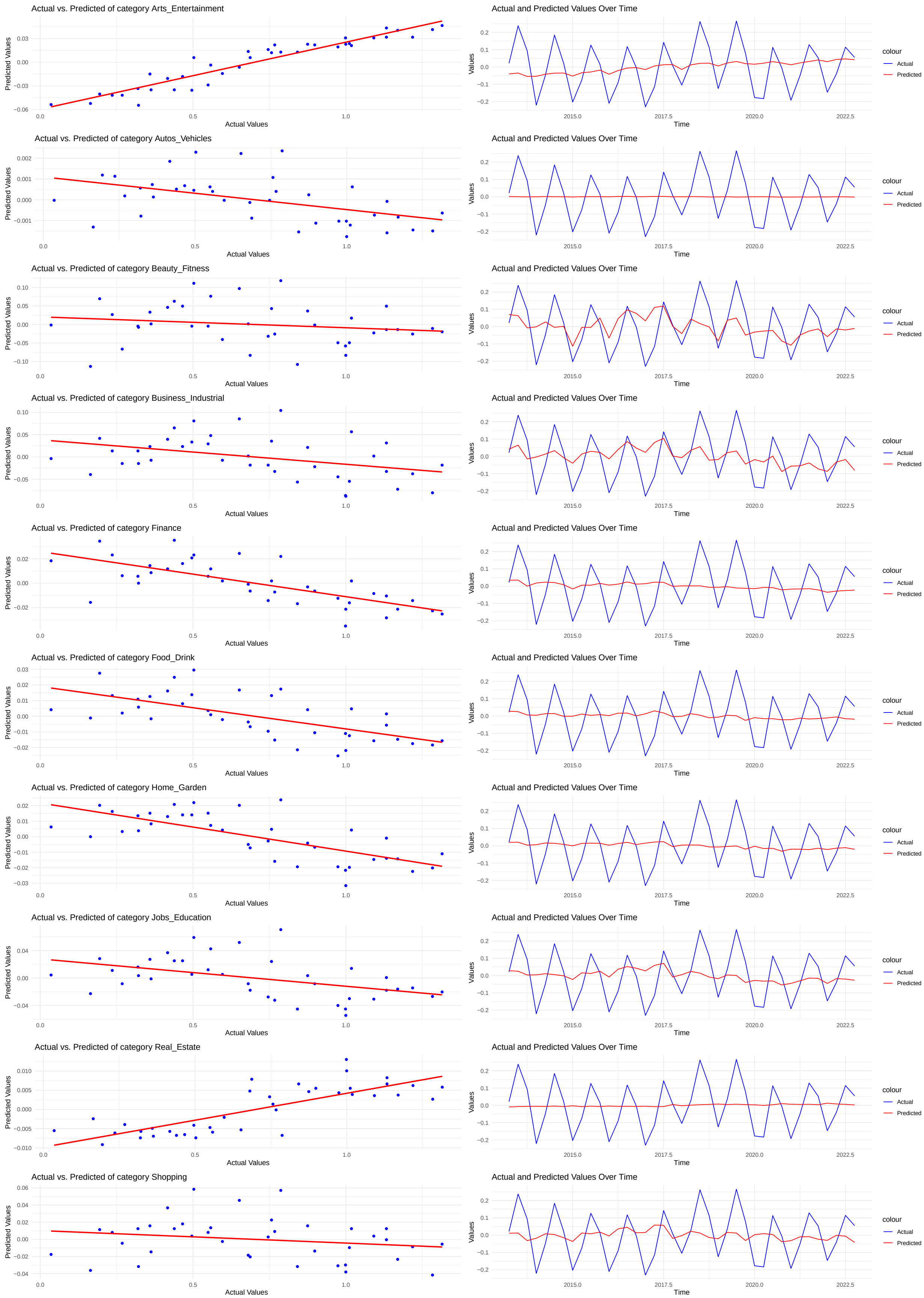


Figure 26: Figure 1: Moldova GDP Over Time

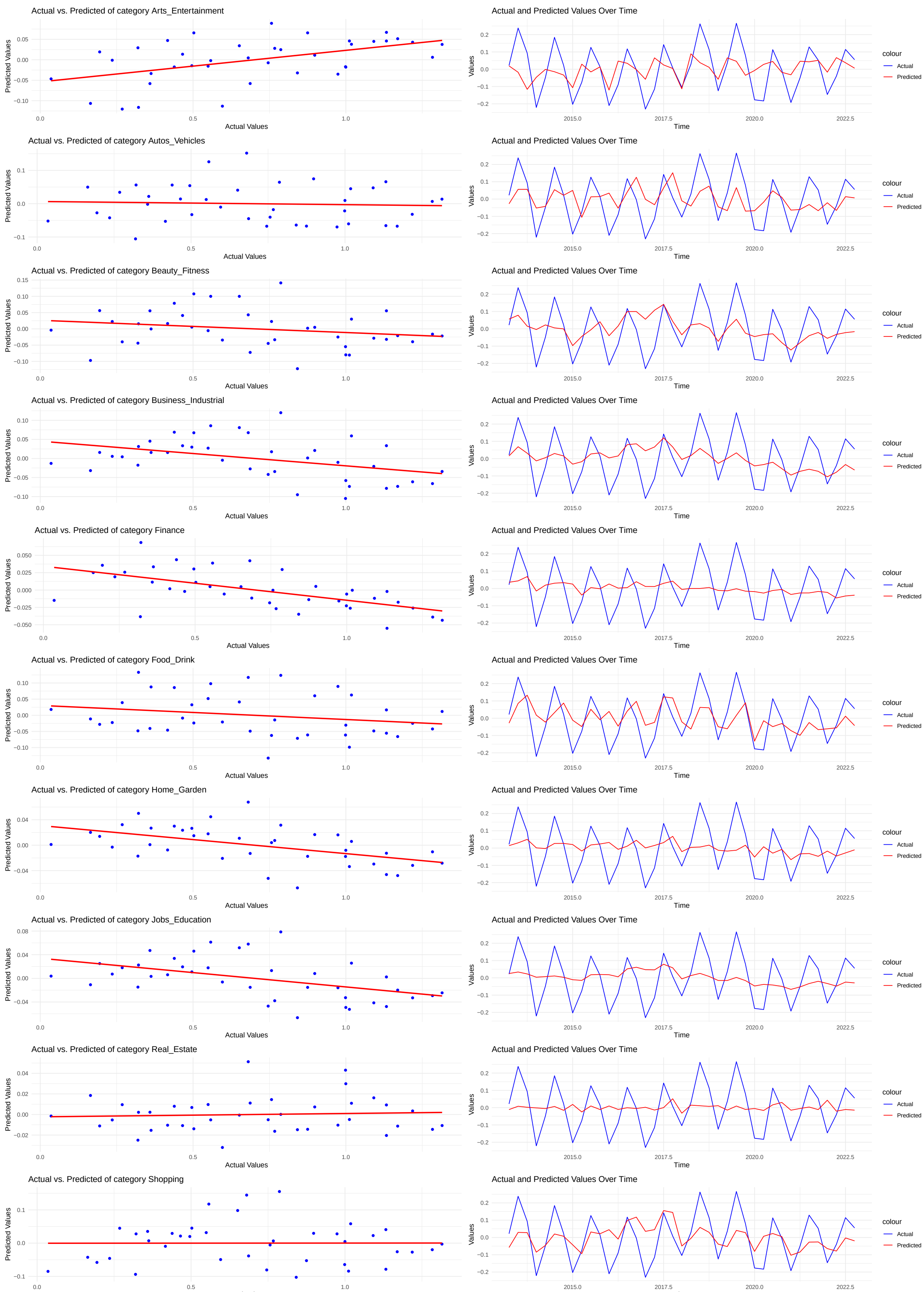
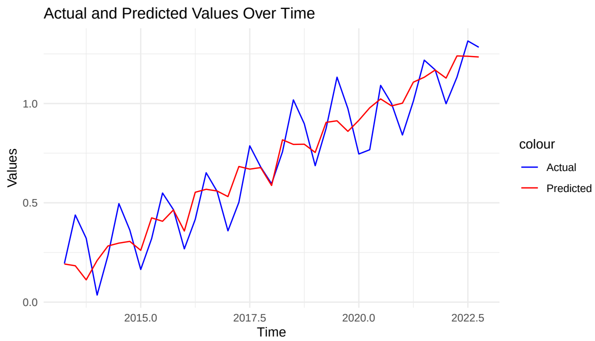
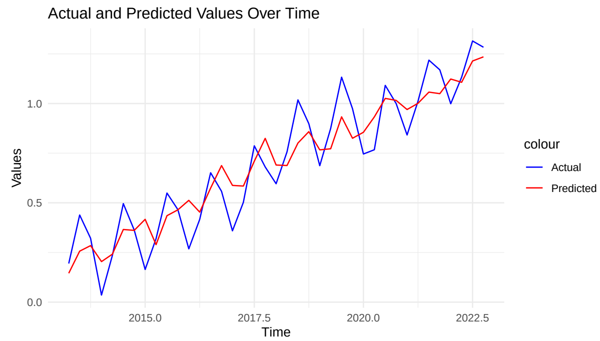
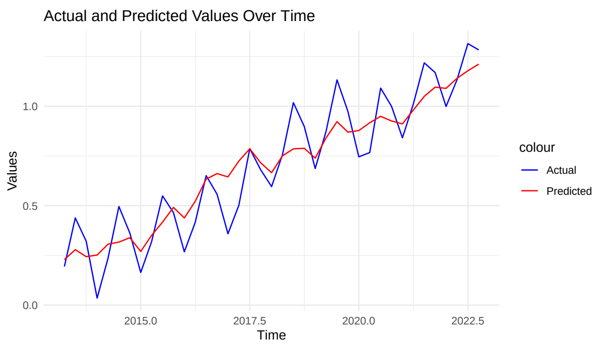
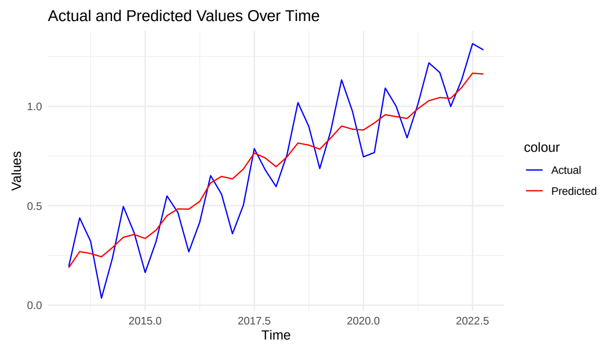
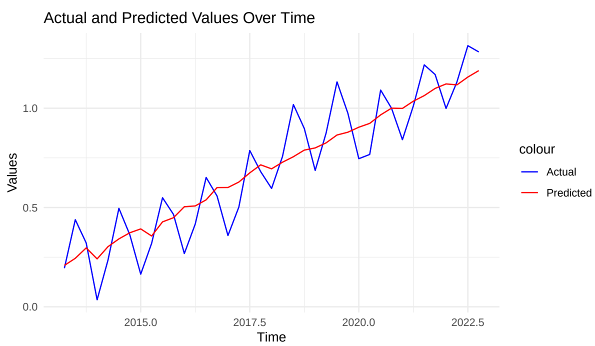
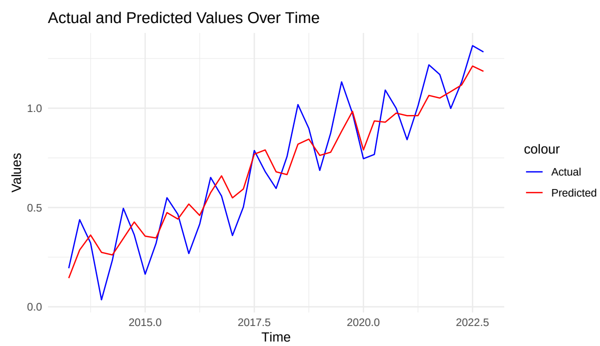
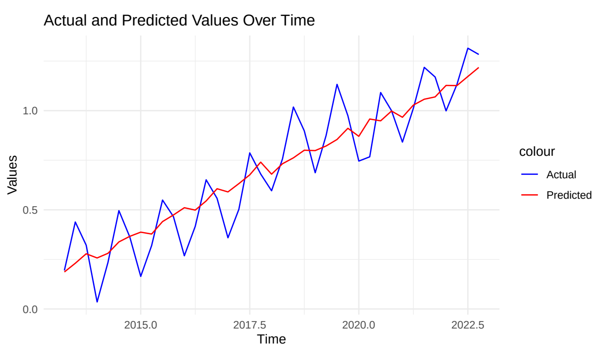
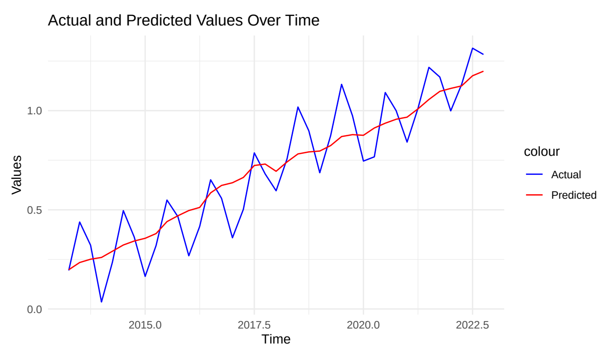
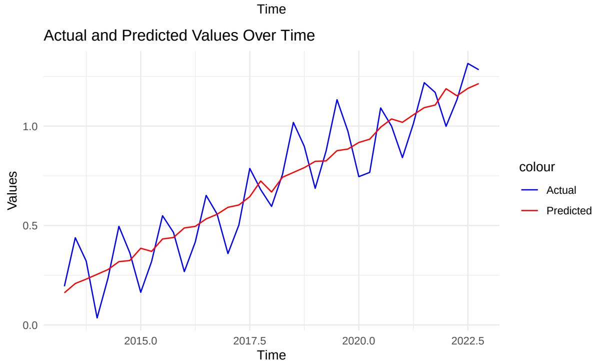
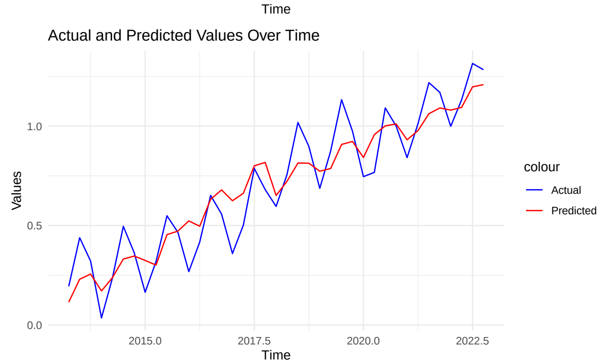
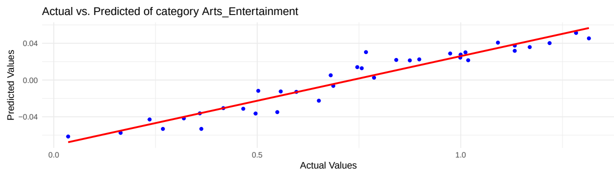
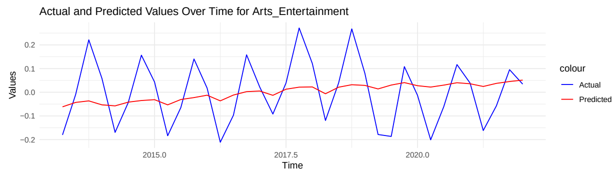
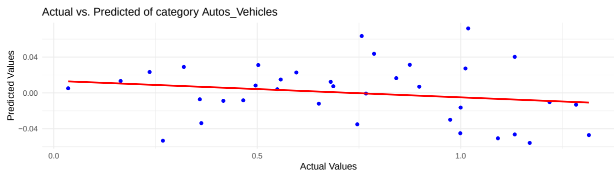
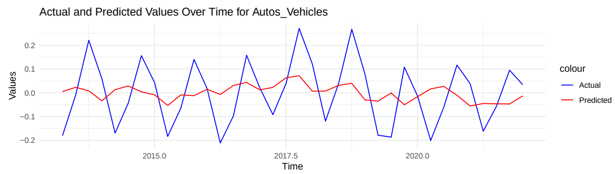
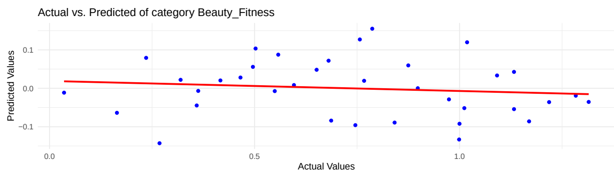
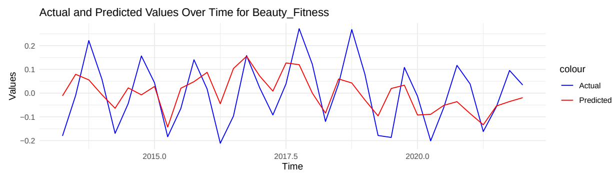
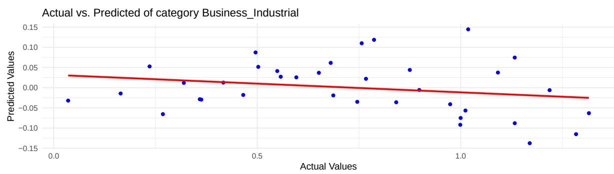
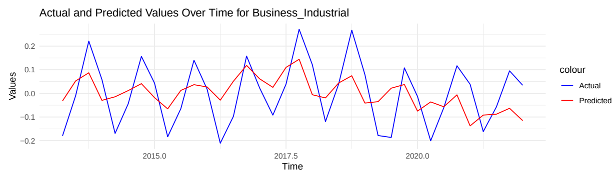
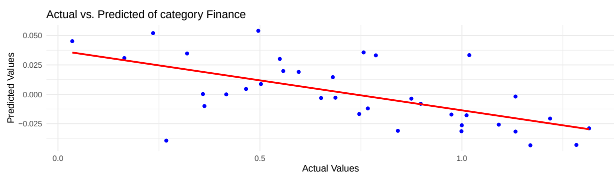
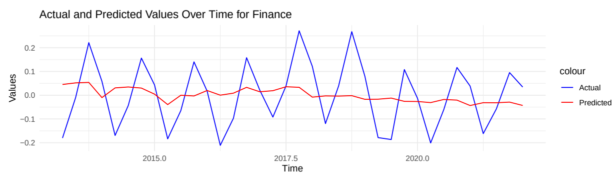
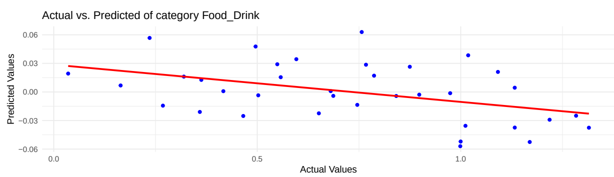
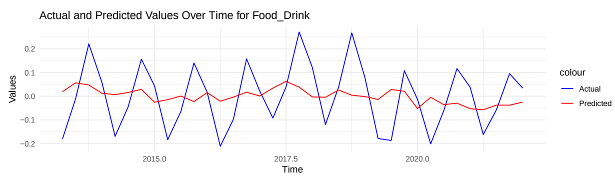
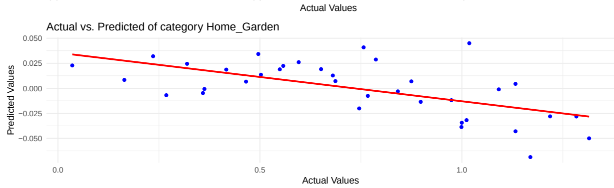
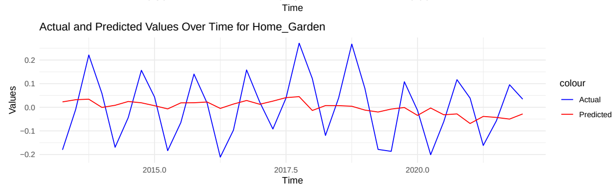
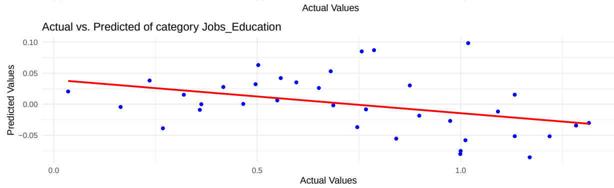
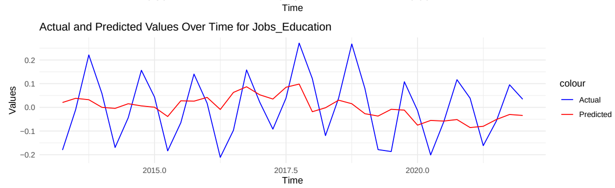
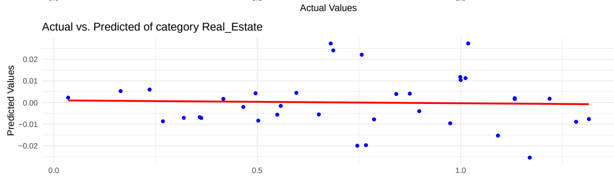
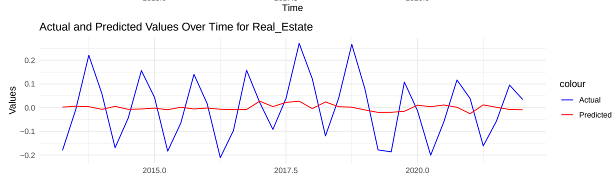
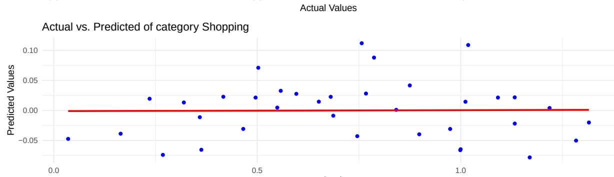


Figure 27: Figure 1: Moldova GDP Over Time


Actual and Predicted Values Over Time
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
2022.5
colour
Actual
Predicted

Actual and Predicted Values Over Time
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
2022.5
colour
Actual
Predicted

Actual and Predicted Values Over Time
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
2022.5
colour
Actual
Predicted

Actual and Predicted Values Over Time
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
2022.5
colour
Actual
Predicted

Actual and Predicted Values Over Time
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
2022.5
colour
Actual
Predicted

Actual and Predicted Values Over Time
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
2022.5
colour
Actual
Predicted

Actual and Predicted Values Over Time
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
2022.5
colour
Actual
Predicted

Actual and Predicted Values Over Time
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
2022.5
colour
Actual
Predicted

Actual and Predicted Values Over Time
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
2022.5
colour
Actual
Predicted

Actual and Predicted Values Over Time
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
2022.5
colour
Actual
Predicted

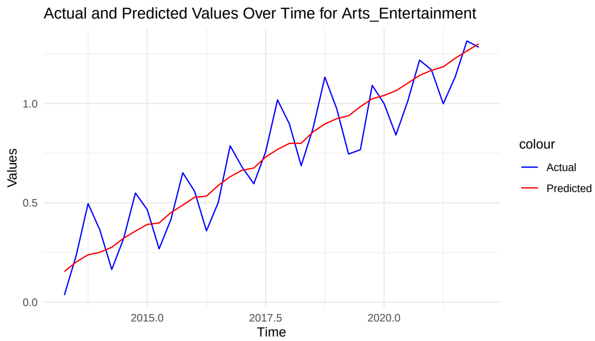
Now we will implement a regressive model, where we fit the linear model $\log GDP_t \sim \log GT_t + \log GT_{t-4}$ model after removing the trend and looking at residuals.

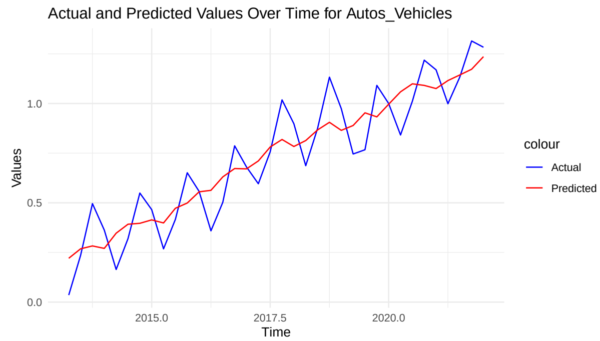
```
## [1] 0.06343166
## [1] 0.05755144
## [1] 0.2967861
## [1] 0.2333422
## [1] 0.04227745
## [1] 0.05026121
## [1] 0.03990783
## [1] 0.1192522
## [1] 0.008369089
## [1] 0.1266087
```

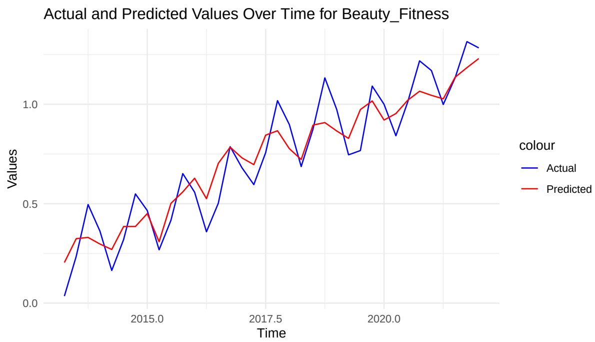

Actual vs. Predicted of category Arts_Entertainment
Predicted Values
Actual Values

Actual and Predicted Values Over Time for Arts_Entertainment
Values
Time
colour
Actual
Predicted

Actual vs. Predicted of category Autos_Vehicles
Predicted Values
Actual Values

Actual and Predicted Values Over Time for Autos_Vehicles
Values
Time
colour
Actual
Predicted

Actual vs. Predicted of category Beauty_Fitness
Predicted Values
Actual Values

Actual and Predicted Values Over Time for Beauty_Fitness
Values
Time
colour
Actual
Predicted

Actual vs. Predicted of category Business_Industrial
Predicted Values
Actual Values

Actual and Predicted Values Over Time for Business_Industrial
Values
Time
colour
Actual
Predicted

Actual vs. Predicted of category Finance
Predicted Values
Actual Values

Actual and Predicted Values Over Time for Finance
Values
Time
colour
Actual
Predicted

Actual vs. Predicted of category Food_Drink
Predicted Values
Actual Values

Actual and Predicted Values Over Time for Food_Drink
Values
Time
colour
Actual
Predicted

Actual vs. Predicted of category Home_Garden
Predicted Values
Actual Values

Actual and Predicted Values Over Time for Home_Garden
Values
Time
colour
Actual
Predicted

Actual vs. Predicted of category Jobs_Education
Predicted Values
Actual Values

Actual and Predicted Values Over Time for Jobs_Education
Values
Time
colour
Actual
Predicted

Actual vs. Predicted of category Real_Estate
Predicted Values
Actual Values

Actual and Predicted Values Over Time for Real_Estate
Values
Time
colour
Actual
Predicted

Actual vs. Predicted of category Shopping
Predicted Values
Actual Values

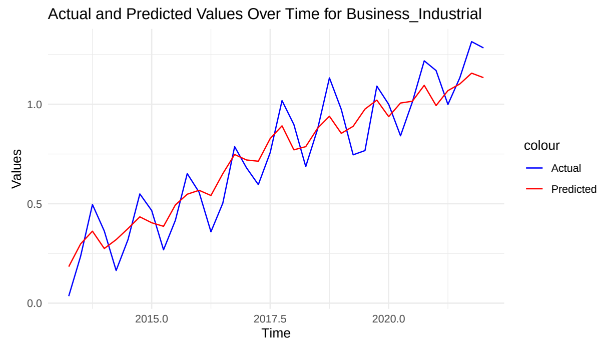
Actual and Predicted Values Over Time for Shopping
Values
Time
colour
Actual
Predicted

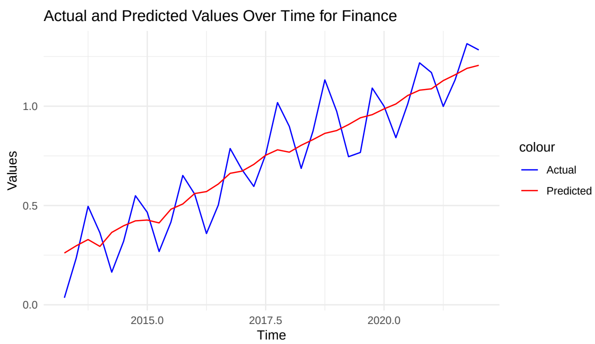
Now we will add the trend back to the predicted values.

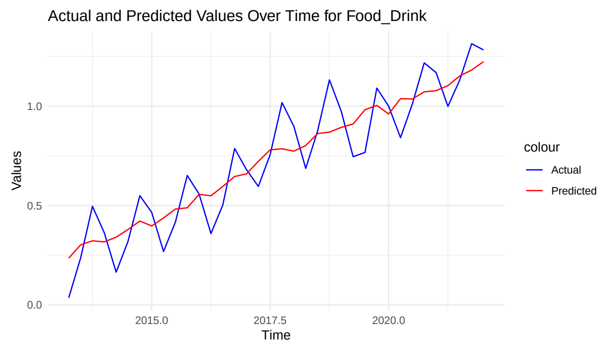

Actual and Predicted Values Over Time for Arts_Entertainment
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
colour
Actual
Predicted

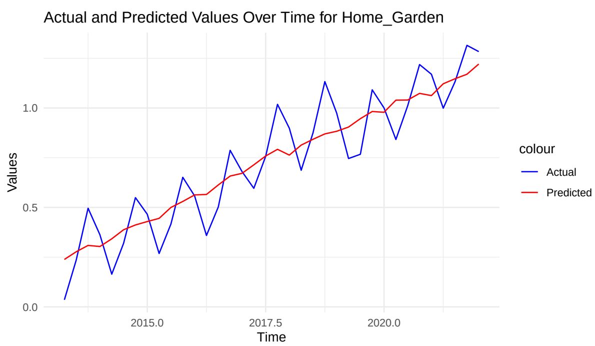

Actual and Predicted Values Over Time for Autos_Vehicles
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
colour
Actual
Predicted

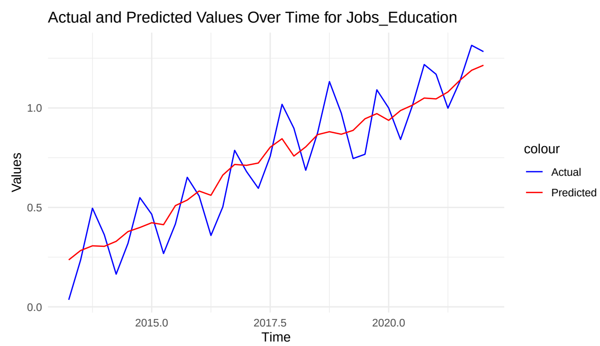

Actual and Predicted Values Over Time for Beauty_Fitness
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
colour
Actual
Predicted

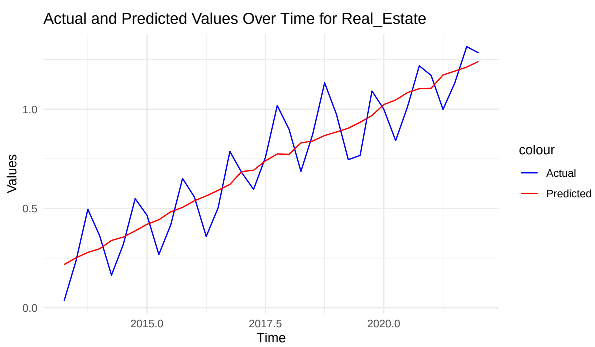

Actual and Predicted Values Over Time for Business_Industrial
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
colour
Actual
Predicted

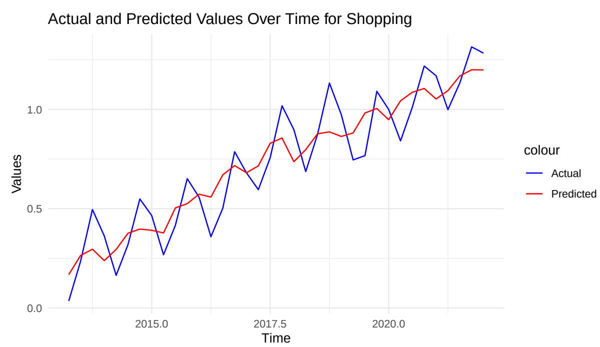

Actual and Predicted Values Over Time for Finance
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
colour
Actual
Predicted


Actual and Predicted Values Over Time for Food_Drink
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
colour
Actual
Predicted


Actual and Predicted Values Over Time for Home_Garden
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
colour
Actual
Predicted


Actual and Predicted Values Over Time for Jobs_Education
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
colour
Actual
Predicted


Actual and Predicted Values Over Time for Real_Estate
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
colour
Actual
Predicted


Actual and Predicted Values Over Time for Shopping
Values
Time
0.0
0.5
1.0
2015.0
2017.5
2020.0
colour
Actual
Predicted

Now we will implement a regressive model, where we fit the linear model $GDP_t \sim GT_t$ model after removing the trend and looking at residuals.

```
## [1] 0.02184263
## [1] 0.00250364
## [1] 0.122836
## [1] 0.06262337
## [1] 0.004689259
## [1] 0.002164315
## [1] 0.008552093
## [1] 0.04213259
## [1] 0.0002931077
## [1] 0.0114733
```

Now we will implement a regressive model, where we fit the linear model $GDP_t \sim GT_t + GT_{t-4}$ model after removing the trend and looking at residuals.

```
## [1] 0.03911929
## [1] 0.01492136
## [1] 0.1818495
## [1] 0.07585063
## [1] 0.01982884
## [1] 0.01218288
## [1] 0.009950445
## [1] 0.05164411
## [1] 0.0004809743
## [1] 0.02890223
```

Now we will implement the $GDP_t \sim GT_t + GT_{t-1} + GT_{t-2} + GT_{t-3} + GT_{t-4}$

Now we will model $GDP_t \sim \sum_i GT_{t,i}$

```
##
## Call:
## lm(formula = formula, data = combined_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11859.9  -2993.3   -779.6   3613.2  11408.0
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         73081.27   21208.00   3.446  0.00176 **
## Arts_Entertainment   -601.93     229.17  -2.627  0.01363 *
## Autos_Vehicles        618.29     352.48   1.754  0.08998 .
## Beauty_Fitness       -621.46     440.47  -1.411  0.16891
## Business_Industrial  -449.44     428.64  -1.049  0.30305
## Finance               453.91     199.50   2.275  0.03046 *
## Food_Drink             66.02     405.45   0.163  0.87178
## Home_Garden           299.46     403.00   0.743  0.46342
## Jobs_Education        -77.78     331.81  -0.234  0.81631
## Real_Estate          -219.18     264.43  -0.829  0.41395
## Shopping              213.35     410.52   0.520  0.60721
```

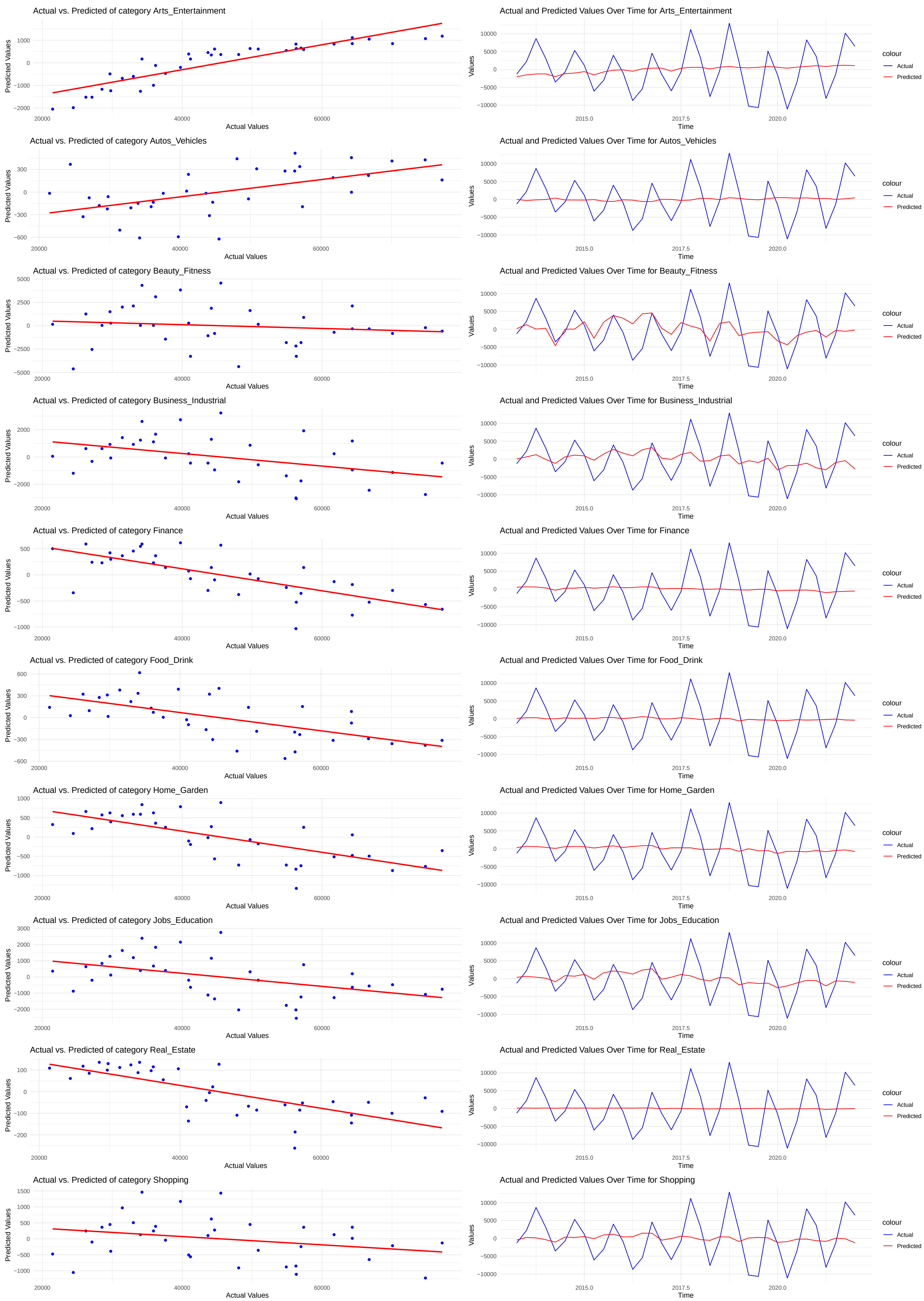


Figure 28: Figure 1: Moldova GDP Over Time

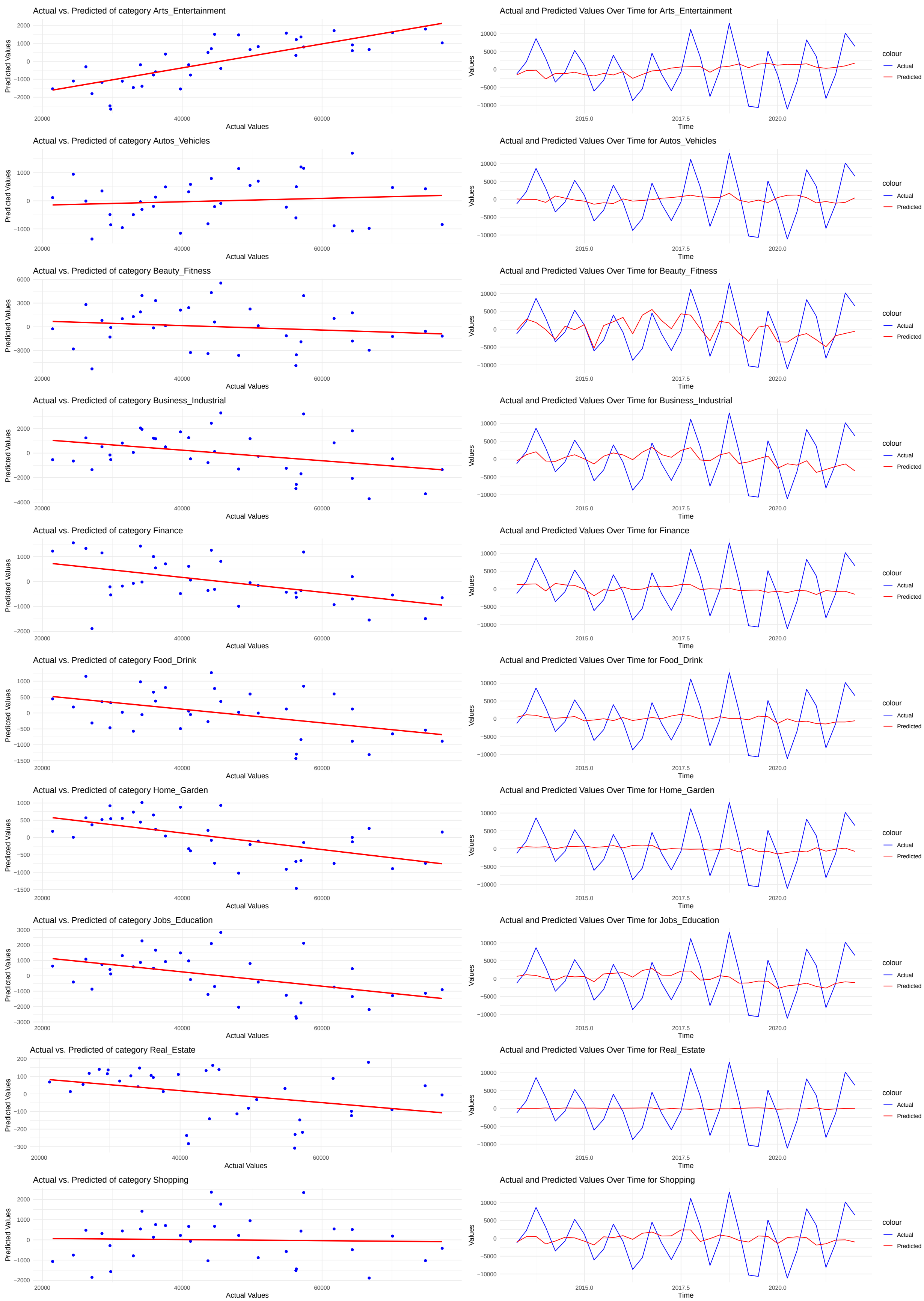


Figure 29: Figure 1: Moldova GDP Over Time

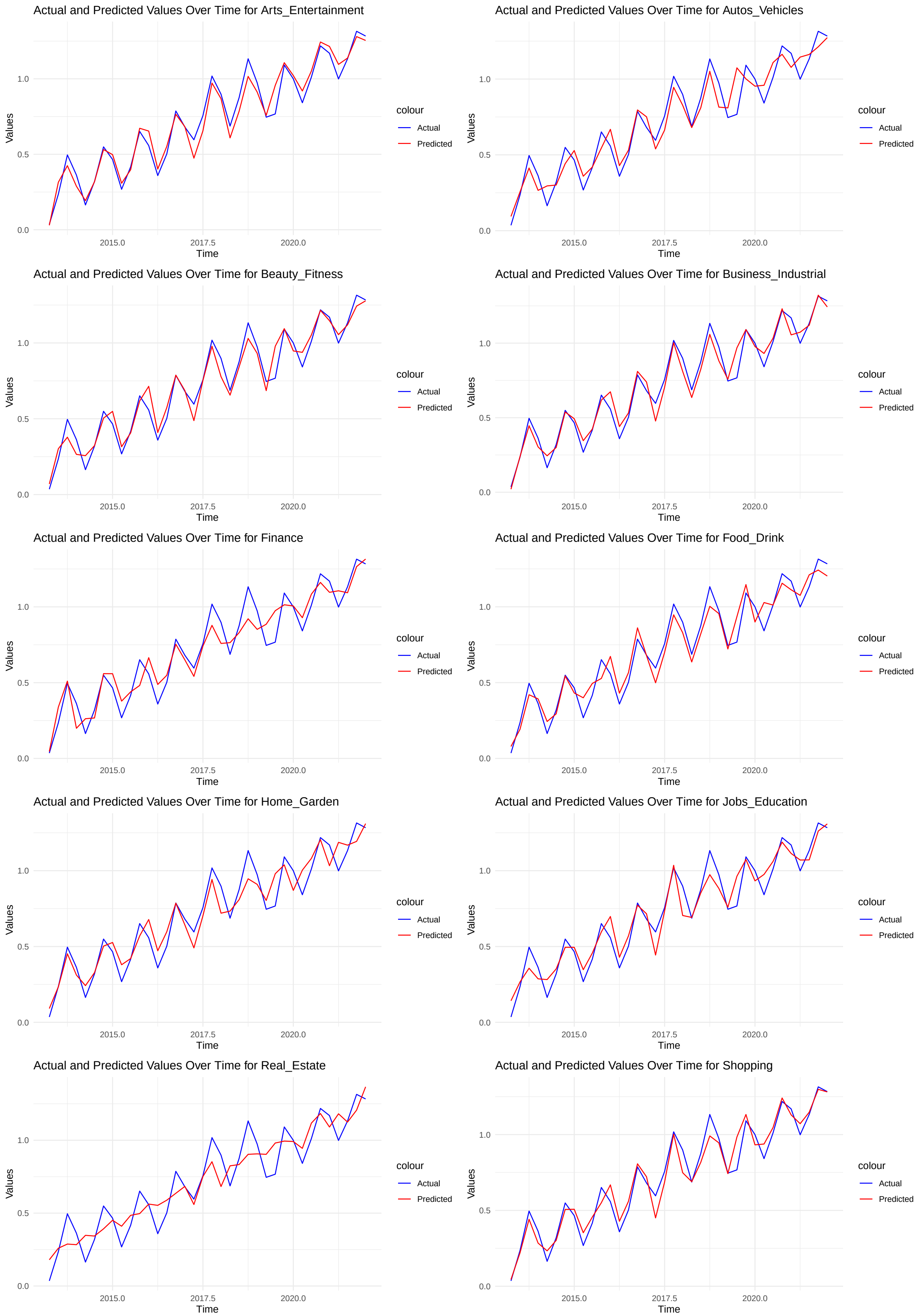


Figure 30: Figure 1: Moldova GDP Over Time

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5759 on 29 degrees of freedom
## Multiple R-squared:  0.897,  Adjusted R-squared:  0.8615
## F-statistic: 25.26 on 10 and 29 DF,  p-value: 1.115e-11
```

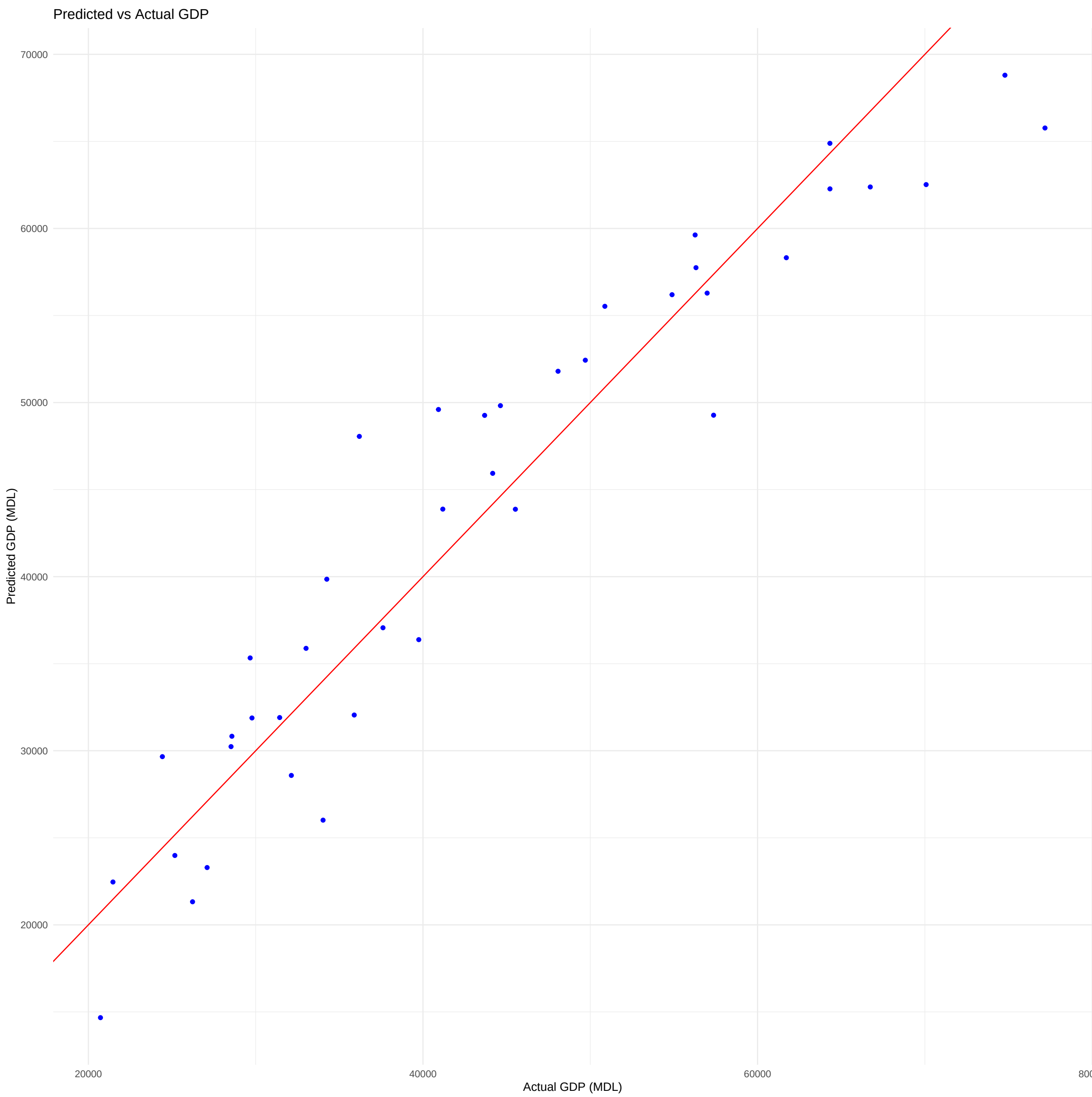
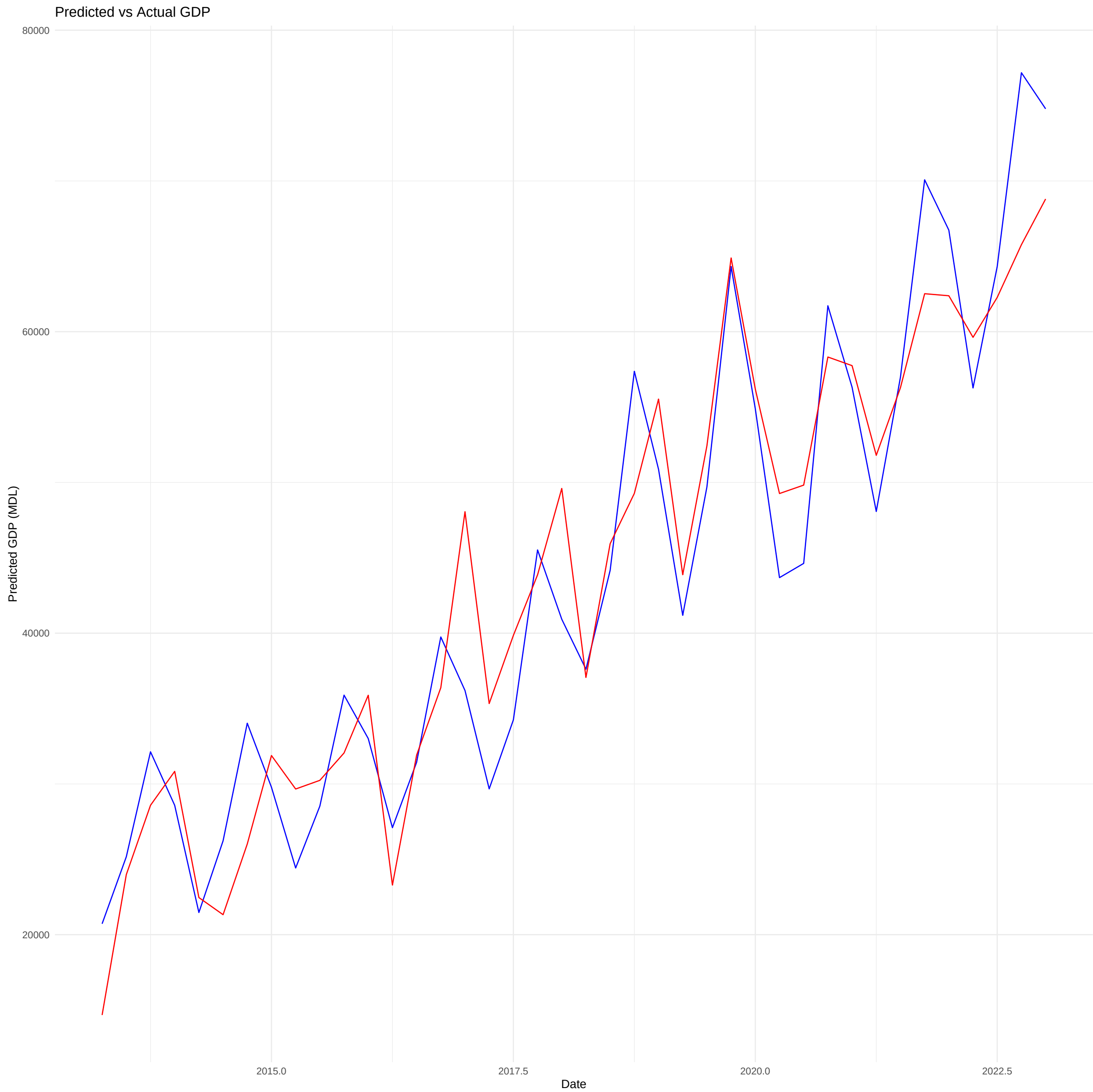


Figure 31: Figure 1: Moldova GDP Over Time



Now a bubble chart

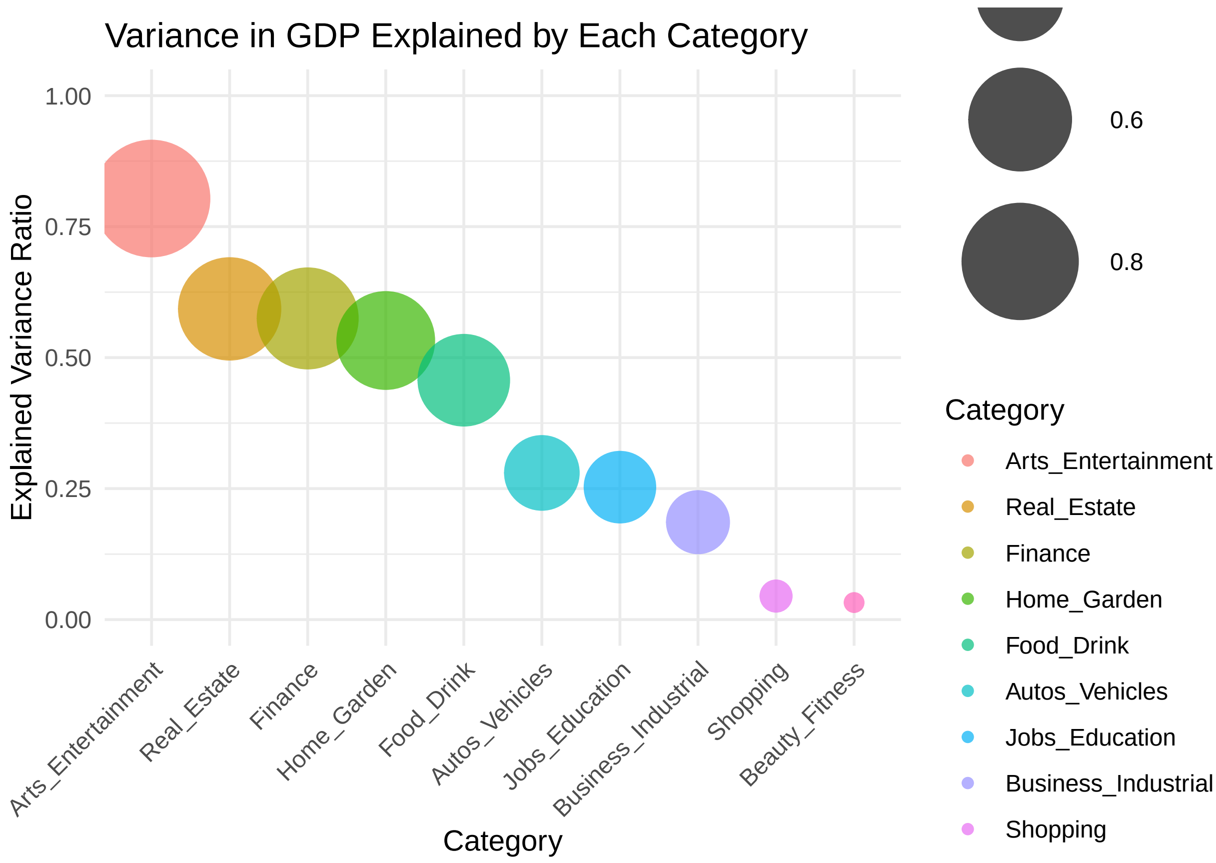


Figure 32: Figure 1: Moldova GDP Over Time