# Untitled

l

```
Installing package into 'C:/Users/vasud/AppData/Local/R/win-library/4.4'
(as 'lib' is unspecified)
package 'ggthemes' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
    C:\Users\vasud\AppData\Local\Temp\RtmpMFYmLY\downloaded_packages
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

Attaching package: 'data.table'


The following objects are masked from 'package:lubridate':

    hour, isoweek, mday, minute, month, quarter, second, wday, week,
    yday, year


The following objects are masked from 'package:dplyr':

    between, first, last


The following object is masked from 'package:purrr':

    transpose
```

```
Attaching package: 'janitor'

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

    chisq.test, fisher.test


Attaching package: 'jsonlite'

The following object is masked from 'package:purrr':

    flatten


Attaching package: 'plotly'

The following object is masked from 'package:httr':

    config

The following object is masked from 'package:ggplot2':

    last_plot

The following object is masked from 'package:stats':

    filter

The following object is masked from 'package:graphics':

    layout

corrplot 0.95 loaded

Loading required package: zoo
```

```
Attaching package: 'zoo'

The following objects are masked from 'package:data.table':

    yearmon, yearqtr

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

    as.Date, as.Date.numeric

Loading required package: lattice

Attaching package: 'caret'

The following object is masked from 'package:httr':

    progress

The following object is masked from 'package:purrr':

    lift


Attaching package: 'MASS'

The following object is masked from 'package:plotly':

    select

The following object is masked from 'package:dplyr':

    select

Loading required package: Matrix

Attaching package: 'Matrix'
```

```
The following objects are masked from 'package:tidyr':

    expand, pack, unpack


Registered S3 method overwritten by 'quantmod':
  method            from
  as.zoo.data.frame zoo


Attaching package: 'forecast'


The following object is masked from 'package:ggpubr':

    gghistogram


Loading required package: car

Loading required package: carData


Attaching package: 'car'


The following object is masked from 'package:dplyr':

    recode


The following object is masked from 'package:purrr':

    some


Loading required package: survival


Attaching package: 'survival'


The following object is masked from 'package:caret':

    cluster
```

```
Loading required package: StanHeaders


rstan version 2.32.6 (Stan version 2.32.2)


For execution on a local, multicore CPU with excess RAM we recommend calling
options(mc.cores = parallel::detectCores()).
To avoid recompilation of unchanged Stan programs, we recommend calling
rstan_options(auto_write = TRUE)
For within-chain threading using `reduce_sum()` or `map_rect()` Stan functions,
change `threads_per_chain` option:
rstan_options(threads_per_chain = 1)


Do not specify '-march=native' in 'LOCAL_CPPFLAGS' or a Makevars file


Attaching package: 'rstan'


The following object is masked from 'package:tidyr':

    extract


Loading required package: Rcpp

Loading 'brms' package (version 2.22.0). Useful instructions
can be found by typing help('brms'). A more detailed introduction
to the package is available through vignette('brms_overview').


Attaching package: 'brms'


The following object is masked from 'package:rstan':

    loo


The following object is masked from 'package:survival':

    kidney
```

```
The following object is masked from 'package:forecast':

    ma


The following object is masked from 'package:lme4':

    ngrps


The following object is masked from 'package:stats':

    ar



Attaching package: 'coda'


The following object is masked from 'package:rstan':

    traceplot


************
Welcome to BayesFactor 0.9.12-4.7. If you have questions, please contact Richard Morey (richarddmorey@gm

Type BFManual() to open the manual.
************

Loading required package: antitrust


Attaching package: 'antitrust'


The following object is masked from 'package:car':

    logit


The following object is masked from 'package:forecast':

    CV
```

```
Attaching package: 'trade'

The following object is masked from 'package:antitrust':

    sim


Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':

    group_rows


Attaching package: 'officer'

The following object is masked from 'package:readxl':

    read_xlsx


Attaching package: 'xgboost'

The following object is masked from 'package:plotly':

    slice

The following object is masked from 'package:dplyr':

    slice

Loaded glmnet 4.1-8

-- Attaching packages ------------------------------------- tidymodels 1.2.0 --

v broom        1.0.7     v rsample      1.2.1
v dials        1.3.0     v tune         1.2.1
```

```
v infer        1.0.7     v workflows    1.1.4
v modeldata    1.4.0     v workflowsets 1.1.0
v parsnip      1.2.1     v yardstick    1.3.1
v recipes      1.1.0

-- Conflicts ---------------------------------------- tidymodels_conflicts() --
x yardstick::accuracy()    masks forecast::accuracy()
x data.table::between()    masks dplyr::between()
x scales::discard()        masks purrr::discard()
x Matrix::expand()         masks tidyr::expand()
x rstan::extract()         masks tidyr::extract()
x plotly::filter()         masks dplyr::filter(), stats::filter()
x data.table::first()      masks dplyr::first()
x recipes::fixed()         masks stringr::fixed()
x jsonlite::flatten()      masks purrr::flatten()
x kableExtra::group_rows() masks dplyr::group_rows()
x dplyr::lag()             masks stats::lag()
x data.table::last()       masks dplyr::last()
x caret::lift()            masks purrr::lift()
x dials::mixture()         masks brms::mixture()
x Matrix::pack()           masks tidyr::pack()
x rsample::populate()      masks Rcpp::populate()
x yardstick::precision()   masks caret::precision()
x yardstick::recall()      masks caret::recall()
x car::recode()            masks dplyr::recode()
x MASS::select()           masks plotly::select(), dplyr::select()
x yardstick::sensitivity() masks caret::sensitivity()
x xgboost::slice()         masks plotly::slice(), dplyr::slice()
x car::some()              masks purrr::some()
x yardstick::spec()        masks readr::spec()
x yardstick::specificity() masks caret::specificity()
x recipes::step()          masks stats::step()
x data.table::transpose()  masks purrr::transpose()
x Matrix::unpack()         masks tidyr::unpack()
x recipes::update()        masks Matrix::update(), stats::update()
* Search for functions across packages at https://www.tidymodels.org/find/

[1] "All required libraries are loaded!"

Registered S3 method overwritten by 'GGally':
  method from
  +.gg   ggplot2


Attaching package: 'dataMaid'

The following object is masked from 'package:recipes':

    check
```

```
The following object is masked from 'package:infer':

    visualize

The following object is masked from 'package:rmarkdown':

    render

The following object is masked from 'package:lme4':

    isSingular

The following object is masked from 'package:dplyr':

    summarize

Rows: 23977 Columns: 6
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (1): CountryName
dbl (4): Year, Value of global merchandise exports as a share of GDP, GDP pe...
lgl (1): World regions according to OWID

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
colnames(Trade)
```

```
[1] "CountryName"
[2] "Year"
[3] "Value of global merchandise exports as a share of GDP"
[4] "GDP per capita"
[5] "Population (historical)"
[6] "World regions according to OWID"
```

```
print(summary(Trade))
```

```
 CountryName             Year
 Length:23977       Min.   :   1
 Class :character   1st Qu.:1877
 Mode  :character   Median :1961
                    Mean   :1894
                    3rd Qu.:1993
                    Max.   :2022

 Value of global merchandise exports as a share of GDP GDP per capita
 Min.   :  0.00                                        Min.   :   295
 1st Qu.: 11.16                                        1st Qu.:  1470
 Median : 19.03                                        Median :  2618
 Mean   : 25.59                                        Mean   :  6870
 3rd Qu.: 31.36                                        3rd Qu.:  7232
 Max.   :779.76                                        Max.   :160051
```

```
 NA's   :10169                                          NA's   :2391
 Population (historical) World regions according to OWID
 Min.   :8.821e+03       Mode:logical
 1st Qu.:1.991e+06       NA's:23977
 Median :5.704e+06
 Mean   :5.799e+07
 3rd Qu.:1.921e+07
 Max.   :8.021e+09
 NA's   :4953
```

```
colnames(CW_Data)
```

```
[1] "CountryName"                 "Year"
[3] "CurrentYearTotalTradeValue"  "PreviousYearTotalTradeValue"
[5] "GrowthPercentage"            "AverageGrowthPercentage"
```

```
print(summary(CW_Data))
```

```
 CountryName             Year      CurrentYearTotalTradeValue
 Length:4554        Min.   :1989   Min.   :3.000e+00
 Class :character   1st Qu.:2000   1st Qu.:1.634e+06
 Mode  :character   Median :2007   Median :1.361e+07
                    Mean   :2007   Mean   :1.623e+08
                    3rd Qu.:2014   3rd Qu.:1.072e+08
                    Max.   :2021   Max.   :5.181e+09
 PreviousYearTotalTradeValue GrowthPercentage   AverageGrowthPercentage
 Min.   :3.000e+00           Min.   : -89.230   Min.   :-34.43
 1st Qu.:1.528e+06           1st Qu.:  -4.685   1st Qu.:  5.81
 Median :1.274e+07           Median :   6.830   Median :  8.31
 Mean   :1.544e+08           Mean   :  12.361   Mean   : 12.36
 3rd Qu.:9.970e+07           3rd Qu.:  18.538   3rd Qu.: 11.73
 Max.   :4.997e+09           Max.   :5511.730   Max.   :478.50
```

```
AnnVal <- read.csv("C:/Users/vasud/OneDrive/Desktop/U-M - ALL/STATS 506/Final Project Data/KAGGLE/archi
```

```
merged_data_full <- left_join(Trade, CW_Data, by = c("CountryName" = "CountryName", "Year" = "Year"))
merged_data_full2 <- left_join(merged_data_full, AnnVal, by = c("CountryName" = "CountryName", "Year" =
```

```
colnames(merged_data_full2)
```

```
 [1] "CountryName"
 [2] "Year"
 [3] "Value of global merchandise exports as a share of GDP"
 [4] "GDP per capita"
 [5] "Population (historical)"
 [6] "World regions according to OWID"
 [7] "CurrentYearTotalTradeValue"
 [8] "PreviousYearTotalTradeValue"
 [9] "GrowthPercentage"
[10] "AverageGrowthPercentage"
[11] "AnnualTradeValue"
```

```
merged_data_full2 <- merged_data_full2[, !(names(merged_data_full2) %in% c("World regions according to (
colnames(merged_data_full2)
```

```
 [1] "CountryName"
 [2] "Year"
 [3] "Value of global merchandise exports as a share of GDP"
 [4] "GDP per capita"
 [5] "Population (historical)"
 [6] "CurrentYearTotalTradeValue"
 [7] "PreviousYearTotalTradeValue"
 [8] "GrowthPercentage"
 [9] "AverageGrowthPercentage"
[10] "AnnualTradeValue"
```

```
TradeData <- merged_data_full2
colnames(TradeData)
```

```
 [1] "CountryName"
 [2] "Year"
 [3] "Value of global merchandise exports as a share of GDP"
 [4] "GDP per capita"
 [5] "Population (historical)"
 [6] "CurrentYearTotalTradeValue"
 [7] "PreviousYearTotalTradeValue"
 [8] "GrowthPercentage"
 [9] "AverageGrowthPercentage"
[10] "AnnualTradeValue"
```

```
summary(TradeData)
```

```
 CountryName             Year
 Length:23977       Min.   :   1
 Class :character   1st Qu.:1877
 Mode  :character   Median :1961
                    Mean   :1894
                    3rd Qu.:1993
                    Max.   :2022

 Value of global merchandise exports as a share of GDP GDP per capita
 Min.   :  0.00                                        Min.   :   295
 1st Qu.: 11.16                                        1st Qu.:  1470
 Median : 19.03                                        Median :  2618
 Mean   : 25.59                                        Mean   :  6870
 3rd Qu.: 31.36                                        3rd Qu.:  7232
 Max.   :779.76                                        Max.   :160051
 NA's   :10169                                         NA's   :2391
 Population (historical) CurrentYearTotalTradeValue PreviousYearTotalTradeValue
 Min.   :8.821e+03       Min.   :1.250e+02          Min.   :1.950e+02
 1st Qu.:1.991e+06       1st Qu.:2.674e+06          1st Qu.:2.480e+06
```

```
 Median :5.704e+06       Median :1.708e+07          Median :1.587e+07
 Mean   :5.799e+07       Mean   :1.571e+08          Mean   :1.497e+08
 3rd Qu.:1.921e+07       3rd Qu.:1.134e+08          3rd Qu.:1.072e+08
 Max.   :8.021e+09       Max.   :5.181e+09          Max.   :4.997e+09
 NA's   :4953            NA's   :20197              NA's   :20197
 GrowthPercentage  AverageGrowthPercentage AnnualTradeValue
 Min.   : -86.12   Min.   :-34.43          Min.   :1.250e+02
 1st Qu.:  -3.95   1st Qu.:  5.81          1st Qu.:2.329e+06
 Median :   7.06   Median :  8.42          Median :1.465e+07
 Mean   :  10.83   Mean   : 10.61          Mean   :1.490e+08
 3rd Qu.:  18.46   3rd Qu.: 11.57          3rd Qu.:1.031e+08
 Max.   :1951.04   Max.   :292.12          Max.   :5.181e+09
 NA's   :20197     NA's   :20197           NA's   :19947
```

```
summary(TradeData$AnnualTradeValue)
```

```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.      NA's
1.250e+02 2.329e+06 1.465e+07 1.490e+08 1.031e+08 5.181e+09     19947
```

```
TradeData_clean <- na.omit(TradeData)
```

### Cleaning and preparing the data

```
# Remove rows with NA values in specified columns
TradeData <- TradeData %>%
  filter(!is.na(`GDP per capita`) &
         !is.na(`Value of global merchandise exports as a share of GDP`) &
         !is.na(GrowthPercentage))
```

```
summary(TradeData)
```

```
 CountryName             Year
 Length:2680        Min.   :1989
 Class :character   1st Qu.:1998
 Mode  :character   Median :2004
                    Mean   :2003
                    3rd Qu.:2009
                    Max.   :2014
 Value of global merchandise exports as a share of GDP GDP per capita
 Min.   :  1.441                                       Min.   :   561
 1st Qu.: 16.718                                       1st Qu.:  4333
 Median : 25.572                                       Median : 10694
 Mean   : 31.336                                       Mean   : 16165
 3rd Qu.: 40.247                                       3rd Qu.: 24513
 Max.   :183.905                                       Max.   :157713
 Population (historical) CurrentYearTotalTradeValue PreviousYearTotalTradeValue
 Min.   :6.830e+04       Min.   :5.177e+03          Min.   :5.177e+03
 1st Qu.:3.796e+06       1st Qu.:3.210e+06          1st Qu.:2.966e+06
 Median :1.008e+07       Median :1.968e+07          Median :1.769e+07
```

```
 Mean   :4.799e+07       Mean   :1.482e+08          Mean   :1.386e+08
 3rd Qu.:2.922e+07       3rd Qu.:1.161e+08          3rd Qu.:1.067e+08
 Max.   :1.388e+09       Max.   :4.685e+09          Max.   :4.418e+09
 GrowthPercentage   AverageGrowthPercentage AnnualTradeValue
 Min.   : -84.300   Min.   :-34.43          Min.   :5.177e+03
 1st Qu.:  -1.093   1st Qu.:  5.91          1st Qu.:3.210e+06
 Median :   9.205   Median :  8.50          Median :1.968e+07
 Mean   :  12.557   Mean   : 10.33          Mean   :1.482e+08
 3rd Qu.:  20.270   3rd Qu.: 11.70          3rd Qu.:1.161e+08
 Max.   :1951.040   Max.   : 88.73          Max.   :4.685e+09
```

```
library(car)

# Compute VIF
vif(lm(GrowthPercentage ~ `GDP per capita` + `Value of global merchandise exports as a share of GDP`, da
```

```
                                    `GDP per capita`
                                            1.103906
`Value of global merchandise exports as a share of GDP`
                                            1.103906
```

```
# Check correlations between predictors
cor(TradeData[, c("GDP per capita", "Value of global merchandise exports as a share of GDP", "AnnualTrad
```

```
                                                    GDP per capita
GDP per capita                                           1.0000000
Value of global merchandise exports as a share of GDP    0.3067997
AnnualTradeValue                                         0.3809786
                                                    Value of global merchandise exports as a share of
GDP per capita                                                                                   0.3067
Value of global merchandise exports as a share of GDP                                            1.0000
AnnualTradeValue                                                                                 0.0438
                                                    AnnualTradeValue
GDP per capita                                           0.38097857
Value of global merchandise exports as a share of GDP    0.04389992
AnnualTradeValue                                         1.00000000
```

1. **GDP per capita and Annual Trade Value**: The correlation is **0.38**, indicating a moderate positive relationship. This suggests that as GDP per capita increases, the annual trade value tends to increase as well, though not strongly.
2. **Value of Global Merchandise Exports as a Share of GDP and Annual Trade Value**: The correlation is quite **low** at **0.04**, indicating almost no linear relationship between these two variables. This implies that the share of merchandise exports as a portion of GDP is not strongly related to the total annual trade value.
3. **GDP per capita and Value of Global Merchandise Exports as a Share of GDP**: The correlation here is **0.31**, suggesting a weak to moderate positive relationship. This could indicate that richer countries (higher GDP per capita) tend to have a somewhat higher share of exports relative to GDP, but the relationship is not particularly strong.

```r
# Run model with all predictors
lm_full <- lm(GrowthPercentage ~ `GDP per capita` + `Value of global merchandise exports as a share of G

# Compute VIF
vif(lm_full)
```

```
                                       `GDP per capita`
                                               1.297775
`Value of global merchandise exports as a share of GDP`
                                               1.111552
                                      AnnualTradeValue
                                               1.177891
```

```r
# Fit a reduced model (with three variables for VIF)
lm_reduced <- lm(GrowthPercentage ~ `GDP per capita` + `Value of global merchandise exports as a share o

# Compute VIF
vif(lm_reduced)
```

```
                                       `GDP per capita`
                                               1.297775
`Value of global merchandise exports as a share of GDP`
                                               1.111552
                                      AnnualTradeValue
                                               1.177891
```

```r
library(mice)
```

```

Attaching package: 'mice'

The following object is masked from 'package:stats':

    filter

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

    cbind, rbind
```

```r
imputed_data <- mice(TradeData, m = 1, method = 'pmm', maxit = 10, seed = 123)
```

```

 iter imp variable
  1   1
  2   1
  3   1
  4   1
  5   1
  6   1
  7   1
  8   1
```

```
  9   1
  10   1
```

```
Warning: Number of logged events: 2
```

```r
TradeData_imputed <- complete(imputed_data)

# Perform PCA including the new variable
pca <- prcomp(TradeData_imputed[, c("GDP per capita", "Value of global merchandise exports as a share o

# Get the first principal component
TradeData_imputed$PCA1 <- pca$x[, 1]

# Use the first principal component in the regression model
model_pca <- lm(GrowthPercentage ~ PCA1, data = TradeData_imputed)
summary(model_pca)
```

```
Call:
lm(formula = GrowthPercentage ~ PCA1, data = TradeData_imputed)

Residuals:
    Min      1Q  Median      3Q     Max
 -97.77  -13.66   -3.26    7.78 1938.23

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.5574     0.9813  12.797   <2e-16 ***
PCA1         -0.8157     0.7984  -1.022    0.307
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 50.8 on 2678 degrees of freedom
Multiple R-squared:  0.0003896, Adjusted R-squared:  1.635e-05
F-statistic: 1.044 on 1 and 2678 DF,  p-value: 0.307
```

```r
# Run the regression again with the new variable included
model <- lm(GrowthPercentage ~ `GDP per capita` + `Value of global merchandise exports as a share of GD
summary(model)
```

```
Call:
lm(formula = GrowthPercentage ~ `GDP per capita` + `Value of global merchandise exports as a share of G
    AnnualTradeValue, data = TradeData_imputed)

Residuals:
    Min      1Q  Median      3Q     Max
 -96.41  -13.62   -3.18    7.89 1938.90

Coefficients:
```

```
                                                  Estimate Std. Error
(Intercept)                                      1.213e+01  1.787e+00
`GDP per capita`                                -1.058e-04  6.852e-05
`Value of global merchandise exports as a share of GDP`  7.814e-02  4.616e-02
AnnualTradeValue                                -2.124e-09  2.903e-09
                                                t value Pr(>|t|)
(Intercept)                                       6.789 1.38e-11 ***
`GDP per capita`                                 -1.545   0.1225
`Value of global merchandise exports as a share of GDP`   1.693   0.0906 .
AnnualTradeValue                                 -0.732   0.4644
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 50.77 on 2676 degrees of freedom
Multiple R-squared:  0.00221,   Adjusted R-squared:  0.001091
F-statistic: 1.975 on 3 and 2676 DF,  p-value: 0.1155
```

```r
library(glmnet)

# Prepare data with the new column
x <- as.matrix(TradeData[, c("GDP per capita", "Value of global merchandise exports as a share of GDP",
y <- TradeData$GrowthPercentage

# Ridge regression (alpha = 0)
ridge_model <- glmnet(x, y, alpha = 0)

# Lasso regression (alpha = 1)
lasso_model <- glmnet(x, y, alpha = 1)

# Cross-validation to choose lambda
cv_ridge <- cv.glmnet(x, y, alpha = 0)
cv_lasso <- cv.glmnet(x, y, alpha = 1)

# Best lambda values
cv_ridge$lambda.min
```

```
[1] 11.77226
```

```r
cv_lasso$lambda.min
```

```
[1] 0.00890528
```

```r
TradeData_scaled <- TradeData
TradeData_scaled[, c("GDP per capita", "Value of global merchandise exports as a share of GDP", "Annual
```

```r
coef(ridge_model, s = cv_ridge$lambda.min)
```

```
4 x 1 sparse Matrix of class "dgCMatrix"
                                                       s1
(Intercept)                                  1.231165e+01
```

```
GDP per capita                                          -7.984363e-05
Value of global merchandise exports as a share of GDP  5.867669e-02
AnnualTradeValue                                        -2.039838e-09
```

```
coef(lasso_model, s = cv_lasso$lambda.min)
```

```
4 x 1 sparse Matrix of class "dgCMatrix"
                                                                  s1
(Intercept)                                            1.213900e+01
GDP per capita                                         -1.051822e-04
Value of global merchandise exports as a share of GDP  7.758722e-02
AnnualTradeValue                                       -2.109717e-09
```

**Lambda values for Ridge and Lasso:**

- Ridge (**`lambda.min`**): 11.77226
- Lasso (**`lambda.min`**): 0.00890528 These values represent the best regularization parameters selected through cross-validation.

1. **Ridge regression coefficients** at the optimal lambda (**`lambda.min`**):
   - Intercept: 12.31165
   - **`GDP per capita`**: -7.984363e-05
   - **`Value of global merchandise exports as a share of GDP`**: 0.05867669
   - **`AnnualTradeValue`**: -2.039838e-09
2. **Lasso regression coefficients** at the optimal lambda (**`lambda.min`**):
   - Intercept: 12.13900
   - **`GDP per capita`**: -1.051822e-04
   - **`Value of global merchandise exports as a share of GDP`**: 0.07758722
   - **`AnnualTradeValue`**: -2.109717e-09

**Interpretation:**

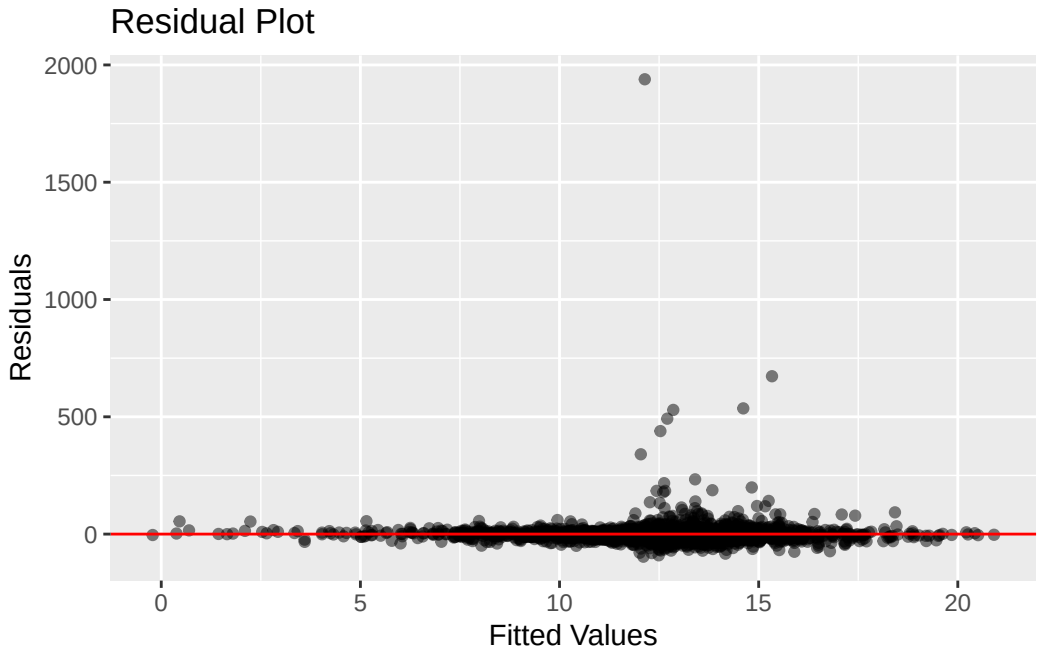
- Both models suggest a negative relationship between **`GDP per capita`** and the dependent variable (**`GrowthPercentage`**), and a positive relationship for the share of global merchandise exports and growth percentage.
- The coefficients for **`AnnualTradeValue`** in both models are very close to zero, suggesting that **`AnnualTradeValue`** has a minimal impact on the predicted **`GrowthPercentage`** after regularization

At this point, I want to check if adding the Annual Trade Value helped my model:

```
# Run a full linear model with all predictors including the new column
model_full <- lm(GrowthPercentage ~ `GDP per capita` + `Value of global merchandise exports as a share

# Summary of the model to check adjusted R-squared and other diagnostics
summary(model_full)
```

```
Call:
lm(formula = GrowthPercentage ~ `GDP per capita` + `Value of global merchandise exports as a share of GD
    AnnualTradeValue, data = TradeData)

Residuals:
    Min      1Q  Median      3Q     Max
 -96.41  -13.62   -3.18    7.89 1938.90

Coefficients:
                                                    Estimate Std. Error
(Intercept)                                        1.213e+01  1.787e+00
`GDP per capita`                                  -1.058e-04  6.852e-05
`Value of global merchandise exports as a share of GDP`  7.814e-02  4.616e-02
AnnualTradeValue                                  -2.124e-09  2.903e-09
                                                  t value Pr(>|t|)
(Intercept)                                         6.789 1.38e-11 ***
`GDP per capita`                                   -1.545   0.1225
`Value of global merchandise exports as a share of GDP`   1.693   0.0906 .
AnnualTradeValue                                   -0.732   0.4644
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 50.77 on 2676 degrees of freedom
Multiple R-squared:  0.00221,   Adjusted R-squared:  0.001091
F-statistic: 1.975 on 3 and 2676 DF,  p-value: 0.1155
```

```r
# Plot residuals to assess if the new predictor improves the model
ggplot(TradeData, aes(x = fitted(model_full), y = residuals(model_full))) +
  geom_point(alpha = 0.5) +
  geom_hline(yintercept = 0, color = "red") +
  labs(
    title = "Residual Plot",
    x = "Fitted Values",
    y = "Residuals"
  )
```




```
# Compare models by adjusted R-squared
model_no_new <- lm(GrowthPercentage ~ `GDP per capita` + `Value of global merchandise exports as a share
summary(model_no_new)$adj.r.squared
```

```
[1] 0.001264414
```

```
summary(model_full)$adj.r.squared
```

```
[1] 0.001091052
```

```
# Check for improvements in the model
if (summary(model_full)$adj.r.squared > summary(model_no_new)$adj.r.squared) {
  print("The new predictor improved the model.")
} else {
  print("The new predictor did not improve the model.")
}
```

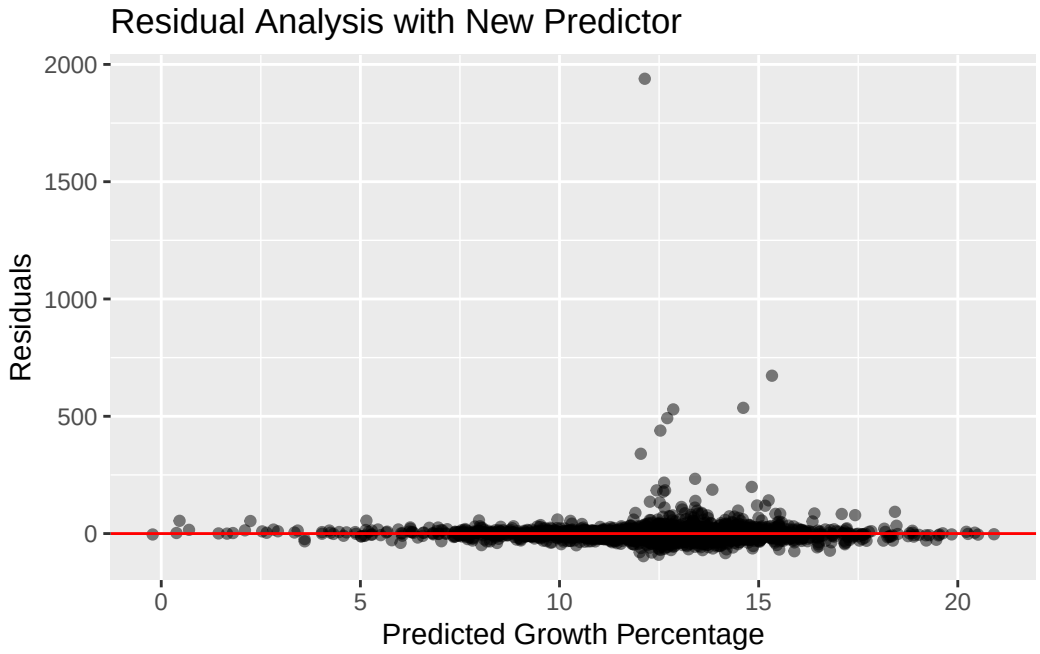
```
[1] "The new predictor did not improve the model."
```

The new predictor did not necessarily improve the model.

***We can now do some predictive model analysis***

```
# Run a linear model with all predictors including the new column
linear_model_full <- lm(GrowthPercentage ~ `GDP per capita` + `Value of global merchandise exports as a

# Add predictions to dataset
TradeData <- TradeData %>%
  mutate(PredictedGrowth = predict(linear_model_full, newdata = TradeData))

# Residual Analysis
ggplot(TradeData, aes(x = PredictedGrowth, y = GrowthPercentage - PredictedGrowth)) +
  geom_point(alpha = 0.5) +
```

```
  geom_hline(yintercept = 0, color = "red") +
  labs(
    title = "Residual Analysis with New Predictor",
    x = "Predicted Growth Percentage",
    y = "Residuals"
  )
```




```
model_1 <- model_no_new
model_2 <- model_full

# Model Summaries
summary(model_1)
```

```
Call:
lm(formula = GrowthPercentage ~ `GDP per capita` + `Value of global merchandise exports as a share of GI
    data = TradeData)

Residuals:
    Min      1Q  Median      3Q     Max
 -96.22  -13.60   -3.25    7.90 1939.19

Coefficients:
                                                      Estimate Std. Error
(Intercept)                                          1.205e+01  1.783e+00
`GDP per capita`                                    -1.252e-04  6.319e-05
`Value of global merchandise exports as a share of GDP`  8.094e-02  4.599e-02
                                                    t value Pr(>|t|)
(Intercept)                                           6.756 1.74e-11 ***
```

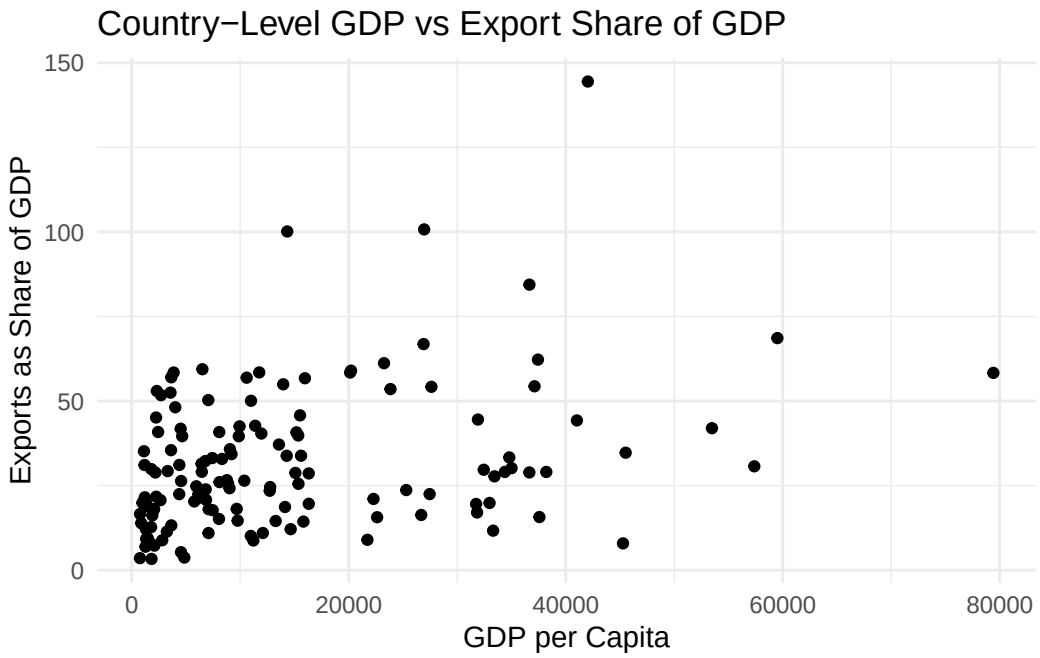
```
`GDP per capita`                                          -1.982   0.0476 *
`Value of global merchandise exports as a share of GDP`   1.760   0.0785 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 50.77 on 2677 degrees of freedom
Multiple R-squared:  0.00201,   Adjusted R-squared:  0.001264
F-statistic: 2.696 on 2 and 2677 DF,  p-value: 0.06767
```

```r
summary(model_2)
```

```
Call:
lm(formula = GrowthPercentage ~ `GDP per capita` + `Value of global merchandise exports as a share of GD
    AnnualTradeValue, data = TradeData)

Residuals:
    Min      1Q  Median      3Q     Max
 -96.41  -13.62   -3.18    7.89 1938.90

Coefficients:
                                                          Estimate Std. Error
(Intercept)                                             1.213e+01  1.787e+00
`GDP per capita`                                       -1.058e-04  6.852e-05
`Value of global merchandise exports as a share of GDP`  7.814e-02  4.616e-02
AnnualTradeValue                                       -2.124e-09  2.903e-09
                                                       t value Pr(>|t|)
(Intercept)                                              6.789 1.38e-11 ***
`GDP per capita`                                        -1.545   0.1225
`Value of global merchandise exports as a share of GDP`   1.693   0.0906 .
AnnualTradeValue                                        -0.732   0.4644
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 50.77 on 2676 degrees of freedom
Multiple R-squared:  0.00221,   Adjusted R-squared:  0.001091
F-statistic: 1.975 on 3 and 2676 DF,  p-value: 0.1155
```

```r
TradeData_country_summary <- TradeData %>%
  group_by(CountryName) %>%
  summarise(mean_gdp_per_capita = mean(`GDP per capita`, na.rm = TRUE),
            mean_exports_share_of_GDP = mean(`Value of global merchandise exports as a share of GDP`, na

# EDA Visualizations
# Scatterplot: GDP per Capita vs. Exports
# Exporting model summaries
broom::tidy(model_1) %>% write.csv("Model_1_Summary.csv")
broom::tidy(model_2) %>% write.csv("Model_2_Summary.csv")
```

```
# Visualize GDP per capita vs exports share for different countries
ggplot(TradeData_country_summary, aes(x = mean_gdp_per_capita, y = mean_exports_share_of_GDP)) +
  geom_point() +
  labs(title = "Country-Level GDP vs Export Share of GDP",
       x = "GDP per Capita",
       y = "Exports as Share of GDP") +
  theme_minimal()
```




```
# Further statistical analysis: e.g., correlation between GDP per capita and export share
cor(TradeData_clean$`GDP per capita`,
    TradeData_clean$`Value of global merchandise exports as a share of GDP`,
    use = "complete.obs")
```

```
[1] 0.3067997
```

```
# For time series analysis, group by Year and summarize export share
TradeData_year_summary <- TradeData_clean %>%
  group_by(Year) %>%
  summarise(mean_exports_share_of_GDP = mean(`Value of global merchandise exports as a share of GDP`, na

# Plot export share over the years
ggplot(TradeData_year_summary, aes(x = Year, y = mean_exports_share_of_GDP)) +
  geom_line(color = "red") +
  labs(title = "Global Merchandise Exports as Share of GDP Over Time",
       x = "Year",
       y = "Mean Share of GDP") +
  theme_minimal()
```


```
# Saving cleaned data for further analysis or export
write_csv(TradeData_clean, "cleaned_data.csv")
```

Both models suggest that the predictors (**GDP per capita**, **Value of global merchandise exports as a share of GDP**, and **AnnualTradeValue**) explain very little of the variation in growth percentage, as indicated by the low R-squared values. Only the **GDP per capita** variable in the first model is statistically significant at the 5% level, with a negative relationship to growth percentage. The export share of GDP has a marginal significance (near the 10% level) but does not show a strong effect.

Thus, the models indicate weak explanatory power and suggest that other factors not included in the model may have a stronger influence on growth percentage.

### Advanced Analysis

```
# Interaction terms in regression
model_3 <- lm(`Value of global merchandise exports as a share of GDP` ~
              `GDP per capita` * GrowthPercentage, data = TradeData)
summary(model_3)
```

```
Call:
lm(formula = `Value of global merchandise exports as a share of GDP` ~
    `GDP per capita` * GrowthPercentage, data = TradeData)

Residuals:
    Min      1Q  Median      3Q     Max
-41.887 -14.014  -5.579   9.059 138.749

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                      2.453e+01  6.207e-01  39.514   <2e-16 ***
```

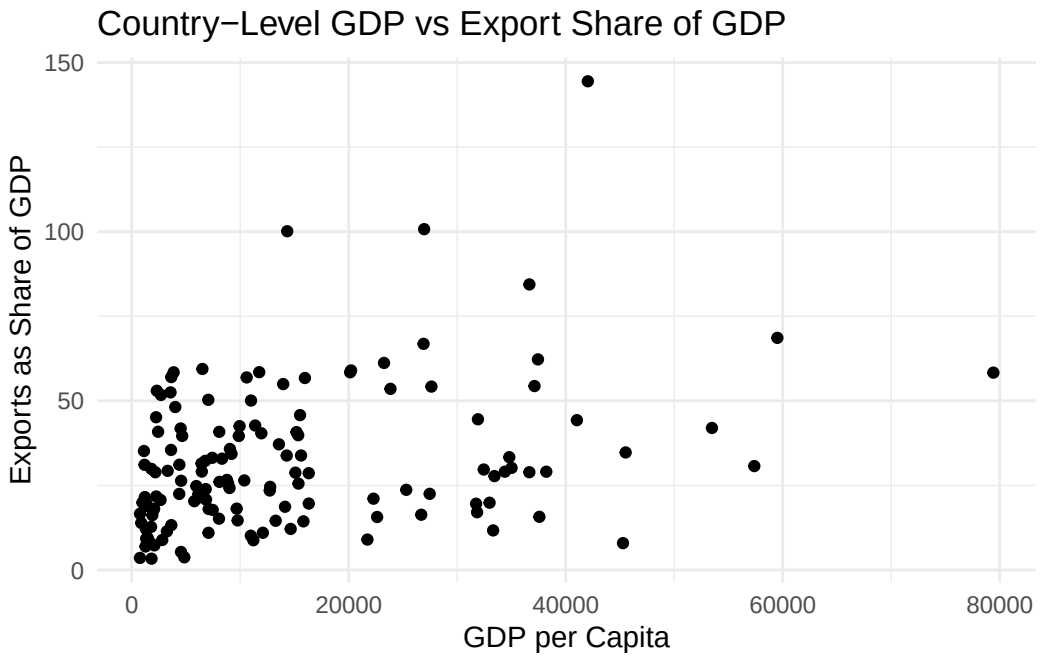
```
`GDP per capita`                  4.097e-04  2.804e-05  14.614   <2e-16 ***
GrowthPercentage                  7.380e-05  1.549e-02   0.005    0.996
`GDP per capita`:GrowthPercentage 1.037e-06  9.633e-07   1.077    0.282
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.32 on 2676 degrees of freedom
Multiple R-squared:  0.09556,   Adjusted R-squared:  0.09455
F-statistic: 94.25 on 3 and 2676 DF,  p-value: < 2.2e-16
```

```r
# Exporting interaction model summary
broom::tidy(model_3) %>% write.csv("Model_3_Summary.csv")

# Feature Engineering
TradeData <- TradeData %>%
  mutate(
    Population_in_millions = TradeData$"Population (historical)" / 1e6,
    GDP_to_population_ratio = `GDP per capita` / TradeData$"Population (historical)"
  )
```

```r
# View the summarized country-level data
head(TradeData_country_summary)
```

```
# A tibble: 6 x 3
  CountryName mean_gdp_per_capita mean_exports_share_of_GDP
  <chr>                     <dbl>                     <dbl>
1 Afghanistan               1828.                      3.33
2 Albania                   7080.                     11.0
3 Algeria                   9055.                     35.8
4 Angola                    7060.                     50.3
5 Argentina                15829.                     14.4
6 Armenia                   8044.                     15.2
```

```r
# Visualize GDP per capita vs exports share for different countries
ggplot(TradeData_country_summary, aes(x = mean_gdp_per_capita, y = mean_exports_share_of_GDP)) +
  geom_point() +
  labs(title = "Country-Level GDP vs Export Share of GDP",
       x = "GDP per Capita",
       y = "Exports as Share of GDP") +
  theme_minimal()
```




```
# Further statistical analysis: e.g., correlation between GDP per capita and export share
cor(TradeData_clean$`GDP per capita`,
    TradeData_clean$`Value of global merchandise exports as a share of GDP`,
    use = "complete.obs")
```

```
[1] 0.3067997
```

```
# For time series analysis, group by Year and summarize export share
TradeData_year_summary <- TradeData_clean %>%
  group_by(Year) %>%
  summarise(mean_exports_share_of_GDP = mean(`Value of global merchandise exports as a share of GDP`, na

# Plot export share over the years
ggplot(TradeData_year_summary, aes(x = Year, y = mean_exports_share_of_GDP)) +
  geom_line(color = "red") +
  labs(title = "Global Merchandise Exports as Share of GDP Over Time",
       x = "Year",
       y = "Mean Share of GDP") +
  theme_minimal()
```


```
# Saving cleaned data for further analysis or export
write_csv(TradeData_clean, "cleaned_data.csv")
```