FDA_Project

Sanskriti_Aath_Yogi

2024-11-05

```
#install.packages("forecast")
```

```
# Load the dataset
data <- read.csv("D:/Aathmika Vijay/7th semester/foundations of data analytics/Gender Inequality Index.
csv")
head(data,6)
```

```
##   ISO3              Country Continent          Hemisphere
## 1  AFG          Afghanistan      Asia Northern Hemisphere
## 2  AGO               Angola    Africa Southern Hemisphere
## 3  ALB              Albania    Europe Northern Hemisphere
## 4  AND              Andorra    Europe Northern Hemisphere
## 5  ARE United Arab Emirates      Asia Northern Hemisphere
## 6  ARG            Argentina   America Southern Hemisphere
##   Human.Development.Groups UNDP.Developing.Regions HDI.Rank..2021.
## 1                      Low                      SA             180
## 2                   Medium                     SSA             148
## 3                     High                     ECA              67
## 4                Very High                                      40
## 5                Very High                      AS              26
## 6                Very High                     LAC              47
##   GII.Rank..2021. Gender.Inequality.Index..1990. Gender.Inequality.Index..1991.
## 1             167                             NA                             NA
## 2             136                          0.725                          0.723
## 3              39                             NA                             NA
## 4              NA                             NA                             NA
## 5              11                          0.659                          0.647
## 6              69                          0.442                          0.439
##   Gender.Inequality.Index..1992. Gender.Inequality.Index..1993.
## 1                             NA                             NA
## 2                          0.721                          0.719
## 3                             NA                             NA
## 4                             NA                             NA
## 5                          0.640                          0.632
## 6                          0.436                          0.430
##   Gender.Inequality.Index..1994. Gender.Inequality.Index..1995.
## 1                             NA                             NA
## 2                          0.717                          0.714
## 3                             NA                             NA
## 4                             NA                             NA
## 5                          0.635                          0.627
## 6                          0.429                          0.426
##   Gender.Inequality.Index..1996. Gender.Inequality.Index..1997.
## 1                             NA                             NA
## 2                          0.712                          0.708
## 3                             NA                             NA
## 4                             NA                             NA
## 5                          0.624                          0.616
## 6                          0.422                          0.419
##   Gender.Inequality.Index..1998. Gender.Inequality.Index..1999.
## 1                             NA                             NA
## 2                          0.683                          0.678
## 3                             NA                          0.330
## 4                             NA                             NA
## 5                          0.609                          0.605
## 6                          0.415                          0.415
##   Gender.Inequality.Index..2000. Gender.Inequality.Index..2001.
## 1                             NA                             NA
## 2                          0.671                          0.665
## 3                          0.319                          0.282
## 4                             NA                             NA
## 5                          0.605                          0.602
## 6                          0.413                          0.412
##   Gender.Inequality.Index..2002. Gender.Inequality.Index..2003.
```

```
## 1                            NA                            NA
## 2                         0.657                         0.649
## 3                         0.297                         0.305
## 4                            NA                            NA
## 5                         0.599                         0.597
## 6                         0.390                         0.385
##   Gender.Inequality.Index..2004. Gender.Inequality.Index..2005.
## 1                            NA                         0.748
## 2                         0.643                         0.636
## 3                         0.301                         0.306
## 4                            NA                            NA
## 5                         0.592                         0.584
## 6                         0.377                         0.373
##   Gender.Inequality.Index..2006. Gender.Inequality.Index..2007.
## 1                         0.749                         0.752
## 2                         0.628                         0.621
## 3                         0.285                         0.286
## 4                            NA                            NA
## 5                         0.311                         0.202
## 6                         0.369                         0.367
##   Gender.Inequality.Index..2008. Gender.Inequality.Index..2009.
## 1                         0.755                         0.755
## 2                         0.568                         0.560
## 3                         0.293                         0.242
## 4                            NA                            NA
## 5                         0.193                         0.185
## 6                         0.363                         0.369
##   Gender.Inequality.Index..2010. Gender.Inequality.Index..2011.
## 1                         0.753                         0.746
## 2                         0.551                         0.544
## 3                         0.246                         0.252
## 4                            NA                            NA
## 5                         0.175                         0.181
## 6                         0.364                         0.362
##   Gender.Inequality.Index..2012. Gender.Inequality.Index..2013.
## 1                         0.738                         0.728
## 2                         0.545                         0.540
## 3                         0.235                         0.225
## 4                            NA                            NA
## 5                         0.171                         0.161
## 6                         0.360                         0.356
##   Gender.Inequality.Index..2014. Gender.Inequality.Index..2015.
## 1                         0.718                         0.706
## 2                         0.531                         0.530
## 3                         0.219                         0.204
## 4                            NA                            NA
## 5                         0.151                         0.126
## 6                         0.351                         0.345
##   Gender.Inequality.Index..2016. Gender.Inequality.Index..2017.
## 1                         0.692                         0.678
## 2                         0.529                         0.538
## 3                         0.191                         0.170
## 4                            NA                            NA
## 5                         0.118                         0.112
## 6                         0.338                         0.328
##   Gender.Inequality.Index..2018. Gender.Inequality.Index..2019.
## 1                         0.671                         0.665
## 2                         0.537                         0.537
```

```
## 3                    0.164                    0.156
## 4                       NA                       NA
## 5                    0.103                    0.056
## 6                    0.315                    0.306
##   Gender.Inequality.Index..2020. Gender.Inequality.Index..2021.
## 1                    0.674                    0.678
## 2                    0.537                    0.537
## 3                    0.156                    0.144
## 4                       NA                       NA
## 5                    0.050                    0.049
## 6                    0.293                    0.287
```

```
str(data)
```

```
## 'data.frame':    195 obs. of  40 variables:
##  $ ISO3                          : chr  "AFG" "AGO" "ALB" "AND" ...
##  $ Country                       : chr  "Afghanistan" "Angola" "Albania" "Andorra" ...
##  $ Continent                     : chr  "Asia" "Africa" "Europe" "Europe" ...
##  $ Hemisphere                    : chr  "Northern Hemisphere" "Southern Hemisphere" "Northern Hemisp
here" "Northern Hemisphere" ...
##  $ Human.Development.Groups      : chr  "Low" "Medium" "High" "Very High" ...
##  $ UNDP.Developing.Regions       : chr  "SA" "SSA" "ECA" "" ...
##  $ HDI.Rank..2021.               : int  180 148 67 40 26 47 85 71 5 25 ...
##  $ GII.Rank..2021.               : int  167 136 39 NA 11 69 53 NA 19 12 ...
##  $ Gender.Inequality.Index..1990.: num  NA 0.725 NA NA 0.659 0.442 0.47 NA 0.185 0.202 ...
##  $ Gender.Inequality.Index..1991.: num  NA 0.723 NA NA 0.647 0.439 0.468 NA 0.184 0.204 ...
##  $ Gender.Inequality.Index..1992.: num  NA 0.721 NA NA 0.64 0.436 0.487 NA 0.184 0.201 ...
##  $ Gender.Inequality.Index..1993.: num  NA 0.719 NA NA 0.632 0.43 0.491 NA 0.181 0.195 ...
##  $ Gender.Inequality.Index..1994.: num  NA 0.717 NA NA 0.635 0.429 0.481 NA 0.179 0.18 ...
##  $ Gender.Inequality.Index..1995.: num  NA 0.714 NA NA 0.627 0.426 0.469 NA 0.176 0.17 ...
##  $ Gender.Inequality.Index..1996.: num  NA 0.712 NA NA 0.624 0.422 0.463 NA 0.174 0.163 ...
##  $ Gender.Inequality.Index..1997.: num  NA 0.708 NA NA 0.616 0.419 0.446 NA 0.172 0.155 ...
##  $ Gender.Inequality.Index..1998.: num  NA 0.683 NA NA 0.609 0.415 0.438 NA 0.169 0.151 ...
##  $ Gender.Inequality.Index..1999.: num  NA 0.678 0.33 NA 0.605 0.415 0.462 NA 0.156 0.147 ...
##  $ Gender.Inequality.Index..2000.: num  NA 0.671 0.319 NA 0.605 0.413 0.453 NA 0.151 0.149 ...
##  $ Gender.Inequality.Index..2001.: num  NA 0.665 0.282 NA 0.602 0.412 0.444 NA 0.15 0.147 ...
##  $ Gender.Inequality.Index..2002.: num  NA 0.657 0.297 NA 0.599 0.39 0.439 NA 0.144 0.145 ...
##  $ Gender.Inequality.Index..2003.: num  NA 0.649 0.305 NA 0.597 0.385 0.408 NA 0.14 0.131 ...
##  $ Gender.Inequality.Index..2004.: num  NA 0.643 0.301 NA 0.592 0.377 0.394 NA 0.136 0.126 ...
##  $ Gender.Inequality.Index..2005.: num  0.748 0.636 0.306 NA 0.584 0.373 0.386 NA 0.135 0.121 ...
##  $ Gender.Inequality.Index..2006.: num  0.749 0.628 0.285 NA 0.311 0.369 0.376 NA 0.134 0.119 ...
##  $ Gender.Inequality.Index..2007.: num  0.752 0.621 0.286 NA 0.202 0.367 0.331 NA 0.134 0.114 ...
##  $ Gender.Inequality.Index..2008.: num  0.755 0.568 0.293 NA 0.193 0.363 0.348 NA 0.138 0.121 ...
##  $ Gender.Inequality.Index..2009.: num  0.755 0.56 0.242 NA 0.185 0.369 0.341 NA 0.135 0.113 ...
##  $ Gender.Inequality.Index..2010.: num  0.753 0.551 0.246 NA 0.175 0.364 0.346 NA 0.136 0.111 ...
##  $ Gender.Inequality.Index..2011.: num  0.746 0.544 0.252 NA 0.181 0.362 0.342 NA 0.131 0.103 ...
##  $ Gender.Inequality.Index..2012.: num  0.738 0.545 0.235 NA 0.171 0.36 0.327 NA 0.13 0.098 ...
##  $ Gender.Inequality.Index..2013.: num  0.728 0.54 0.225 NA 0.161 0.356 0.305 NA 0.123 0.086 ...
##  $ Gender.Inequality.Index..2014.: num  0.718 0.531 0.219 NA 0.151 0.351 0.313 NA 0.117 0.083 ...
##  $ Gender.Inequality.Index..2015.: num  0.706 0.53 0.204 NA 0.126 0.345 0.309 NA 0.111 0.084 ...
##  $ Gender.Inequality.Index..2016.: num  0.692 0.529 0.191 NA 0.118 0.338 0.306 NA 0.101 0.082 ...
##  $ Gender.Inequality.Index..2017.: num  0.678 0.538 0.17 NA 0.112 0.328 0.261 NA 0.096 0.072 ...
##  $ Gender.Inequality.Index..2018.: num  0.671 0.537 0.164 NA 0.103 0.315 0.26 NA 0.092 0.061 ...
##  $ Gender.Inequality.Index..2019.: num  0.665 0.537 0.156 NA 0.056 0.306 0.239 NA 0.08 0.054 ...
##  $ Gender.Inequality.Index..2020.: num  0.674 0.537 0.156 NA 0.05 0.293 0.239 NA 0.074 0.052 ...
##  $ Gender.Inequality.Index..2021.: num  0.678 0.537 0.144 NA 0.049 0.287 0.216 NA 0.073 0.053 ...
```

```
summary(data)
```

```
##      ISO3              Country            Continent          Hemisphere
##  Length:195         Length:195         Length:195         Length:195
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  Human.Development.Groups UNDP.Developing.Regions HDI.Rank..2021.
##  Length:195               Length:195              Min.   :  1.00
##  Class :character         Class :character        1st Qu.: 48.50
##  Mode  :character         Mode  :character        Median : 96.00
##                                                   Mean   : 95.81
##                                                   3rd Qu.:143.50
##                                                   Max.   :191.00
##                                                   NA's   :4
##  GII.Rank..2021.  Gender.Inequality.Index..1990. Gender.Inequality.Index..1991.
##  Min.   :  1.00   Min.   :0.0900                 Min.   :0.0900
##  1st Qu.: 43.25   1st Qu.:0.3195                 1st Qu.:0.3277
##  Median : 85.50   Median :0.5150                 Median :0.5075
##  Mean   : 85.38   Mean   :0.4876                 Mean   :0.4851
##  3rd Qu.:127.75   3rd Qu.:0.6590                 3rd Qu.:0.6495
##  Max.   :170.00   Max.   :0.8110                 Max.   :0.8060
##  NA's   :25       NA's   :67                     NA's   :67
##  Gender.Inequality.Index..1992. Gender.Inequality.Index..1993.
##  Min.   :0.0940                 Min.   :0.0900
##  1st Qu.:0.3292                 1st Qu.:0.3250
##  Median :0.5090                 Median :0.5070
##  Mean   :0.4818                 Mean   :0.4785
##  3rd Qu.:0.6448                 3rd Qu.:0.6438
##  Max.   :0.8030                 Max.   :0.8020
##  NA's   :67                     NA's   :67
##  Gender.Inequality.Index..1994. Gender.Inequality.Index..1995.
##  Min.   :0.0830                 Min.   :0.0750
##  1st Qu.:0.3172                 1st Qu.:0.3145
##  Median :0.5000                 Median :0.4925
##  Mean   :0.4747                 Mean   :0.4706
##  3rd Qu.:0.6385                 3rd Qu.:0.6330
##  Max.   :0.8000                 Max.   :0.7990
##  NA's   :67                     NA's   :65
##  Gender.Inequality.Index..1996. Gender.Inequality.Index..1997.
##  Min.   :0.0670                 Min.   :0.0630
##  1st Qu.:0.3008                 1st Qu.:0.2853
##  Median :0.4855                 Median :0.4795
##  Mean   :0.4656                 Mean   :0.4605
##  3rd Qu.:0.6278                 3rd Qu.:0.6180
##  Max.   :0.7980                 Max.   :0.7960
##  NA's   :65                     NA's   :65
##  Gender.Inequality.Index..1998. Gender.Inequality.Index..1999.
##  Min.   :0.0600                 Min.   :0.0590
##  1st Qu.:0.2980                 1st Qu.:0.2840
##  Median :0.4805                 Median :0.4820
##  Mean   :0.4594                 Mean   :0.4548
##  3rd Qu.:0.6178                 3rd Qu.:0.6240
##  Max.   :0.8120                 Max.   :0.8070
##  NA's   :59                     NA's   :54
##  Gender.Inequality.Index..2000. Gender.Inequality.Index..2001.
```

```
##  Min.   :0.0600                 Min.   :0.0560                
##  1st Qu.:0.2780                 1st Qu.:0.2710                
##  Median :0.4615                 Median :0.4640                
##  Mean   :0.4487                 Mean   :0.4443                
##  3rd Qu.:0.6162                 3rd Qu.:0.6120                
##  Max.   :0.8050                 Max.   :0.8030                
##  NA's   :51                     NA's   :46                    
##  Gender.Inequality.Index..2002. Gender.Inequality.Index..2003.
##  Min.   :0.0540                 Min.   :0.0490                
##  1st Qu.:0.2625                 1st Qu.:0.2610                
##  Median :0.4665                 Median :0.4640                
##  Mean   :0.4387                 Mean   :0.4338                
##  3rd Qu.:0.6050                 3rd Qu.:0.5970                
##  Max.   :0.8020                 Max.   :0.8220                
##  NA's   :45                     NA's   :42                    
##  Gender.Inequality.Index..2004. Gender.Inequality.Index..2005.
##  Min.   :0.0480                 Min.   :0.0470                
##  1st Qu.:0.2567                 1st Qu.:0.2580                
##  Median :0.4670                 Median :0.4540                
##  Mean   :0.4284                 Mean   :0.4232                
##  3rd Qu.:0.5960                 3rd Qu.:0.5930                
##  Max.   :0.8210                 Max.   :0.7990                
##  NA's   :39                     NA's   :36                    
##  Gender.Inequality.Index..2006. Gender.Inequality.Index..2007.
##  Min.   :0.0470                 Min.   :0.0470                
##  1st Qu.:0.2487                 1st Qu.:0.2440                
##  Median :0.4340                 Median :0.4220                
##  Mean   :0.4144                 Mean   :0.4062                
##  3rd Qu.:0.5880                 3rd Qu.:0.5720                
##  Max.   :0.8000                 Max.   :0.7990                
##  NA's   :33                     NA's   :32                    
##  Gender.Inequality.Index..2008. Gender.Inequality.Index..2009.
##  Min.   :0.0480                 Min.   :0.0490                
##  1st Qu.:0.2440                 1st Qu.:0.2320                
##  Median :0.4240                 Median :0.4280                
##  Mean   :0.4039                 Mean   :0.4000                
##  3rd Qu.:0.5650                 3rd Qu.:0.5595                
##  Max.   :0.7980                 Max.   :0.7980                
##  NA's   :32                     NA's   :32                    
##  Gender.Inequality.Index..2010. Gender.Inequality.Index..2011.
##  Min.   :0.0490                 Min.   :0.0440                
##  1st Qu.:0.2310                 1st Qu.:0.2270                
##  Median :0.4220                 Median :0.4240                
##  Mean   :0.3987                 Mean   :0.3939                
##  3rd Qu.:0.5580                 3rd Qu.:0.5555                
##  Max.   :0.7980                 Max.   :0.7990                
##  NA's   :30                     NA's   :28                    
##  Gender.Inequality.Index..2012. Gender.Inequality.Index..2013.
##  Min.   :0.0420                 Min.   :0.0390                
##  1st Qu.:0.2290                 1st Qu.:0.2165                
##  Median :0.4090                 Median :0.3980                
##  Mean   :0.3875                 Mean   :0.3797                
##  3rd Qu.:0.5515                 3rd Qu.:0.5450                
##  Max.   :0.8000                 Max.   :0.8020                
##  NA's   :28                     NA's   :28                    
##  Gender.Inequality.Index..2014. Gender.Inequality.Index..2015.
##  Min.   :0.0350                 Min.   :0.0330                
##  1st Qu.:0.2180                 1st Qu.:0.2115                
```

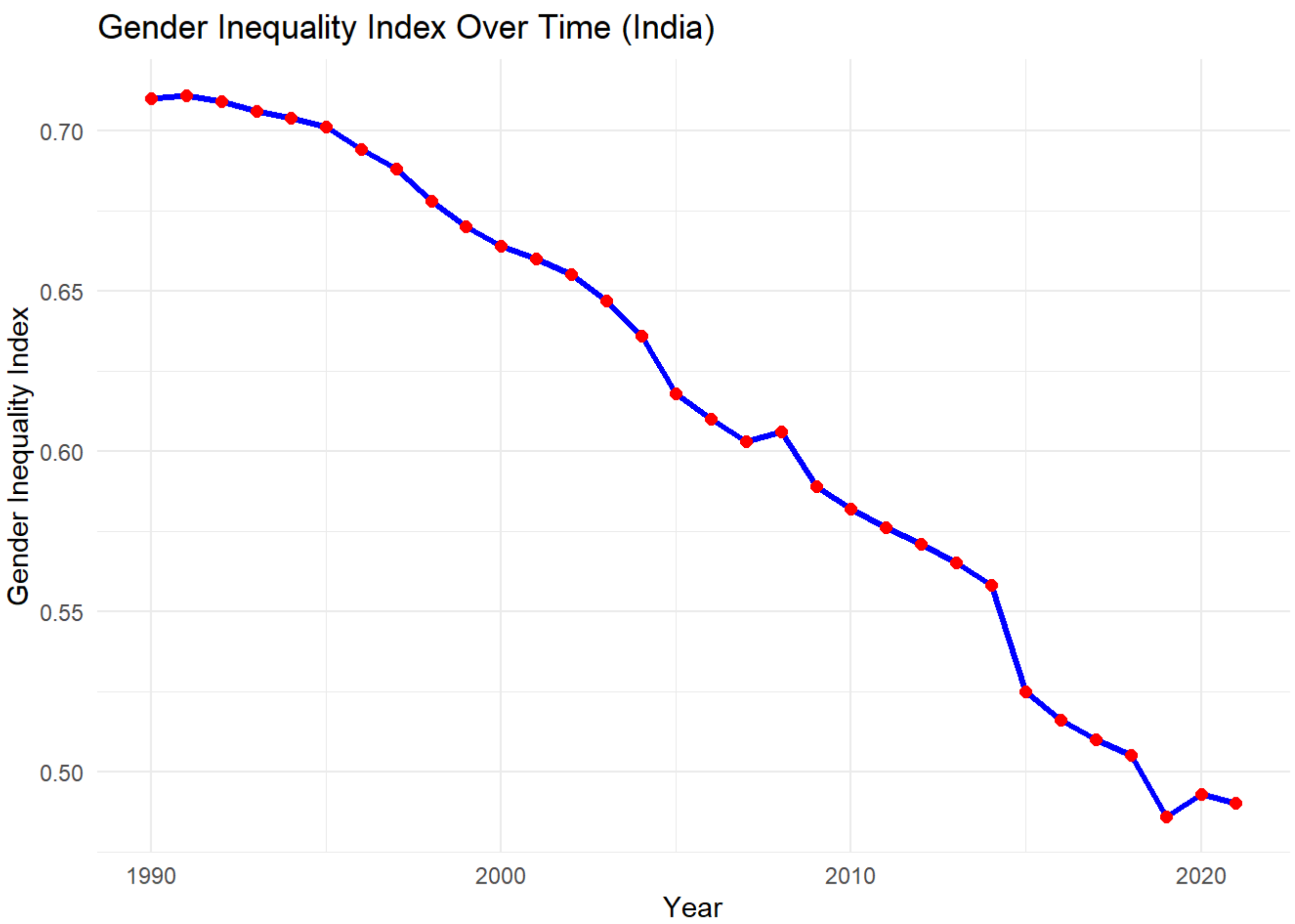
```
##  Median :0.3960                Median :0.3950
##  Mean   :0.3772                Mean   :0.3718
##  3rd Qu.:0.5450                3rd Qu.:0.5327
##  Max.   :0.8060                Max.   :0.8150
##  NA's   :28                    NA's   :27
##  Gender.Inequality.Index..2016. Gender.Inequality.Index..2017.
##  Min.   :0.0300                Min.   :0.0230
##  1st Qu.:0.1920                1st Qu.:0.1893
##  Median :0.3900                Median :0.3780
##  Mean   :0.3652                Mean   :0.3559
##  3rd Qu.:0.5270                3rd Qu.:0.5165
##  Max.   :0.8100                Max.   :0.8080
##  NA's   :26                    NA's   :25
##  Gender.Inequality.Index..2018. Gender.Inequality.Index..2019.
##  Min.   :0.0180                Min.   :0.0160
##  1st Qu.:0.1832                1st Qu.:0.1710
##  Median :0.3785                Median :0.3700
##  Mean   :0.3518                Mean   :0.3469
##  3rd Qu.:0.5162                3rd Qu.:0.5105
##  Max.   :0.8060                Max.   :0.7850
##  NA's   :25                    NA's   :25
##  Gender.Inequality.Index..2020. Gender.Inequality.Index..2021.
##  Min.   :0.0130                Min.   :0.0130
##  1st Qu.:0.1713                1st Qu.:0.1775
##  Median :0.3660                Median :0.3630
##  Mean   :0.3442                Mean   :0.3444
##  3rd Qu.:0.5035                3rd Qu.:0.5058
##  Max.   :0.7840                Max.   :0.8200
##  NA's   :25                    NA's   :25
```

```r
# Load necessary library
#install.packages("ggplot2")
library(ggplot2)
```

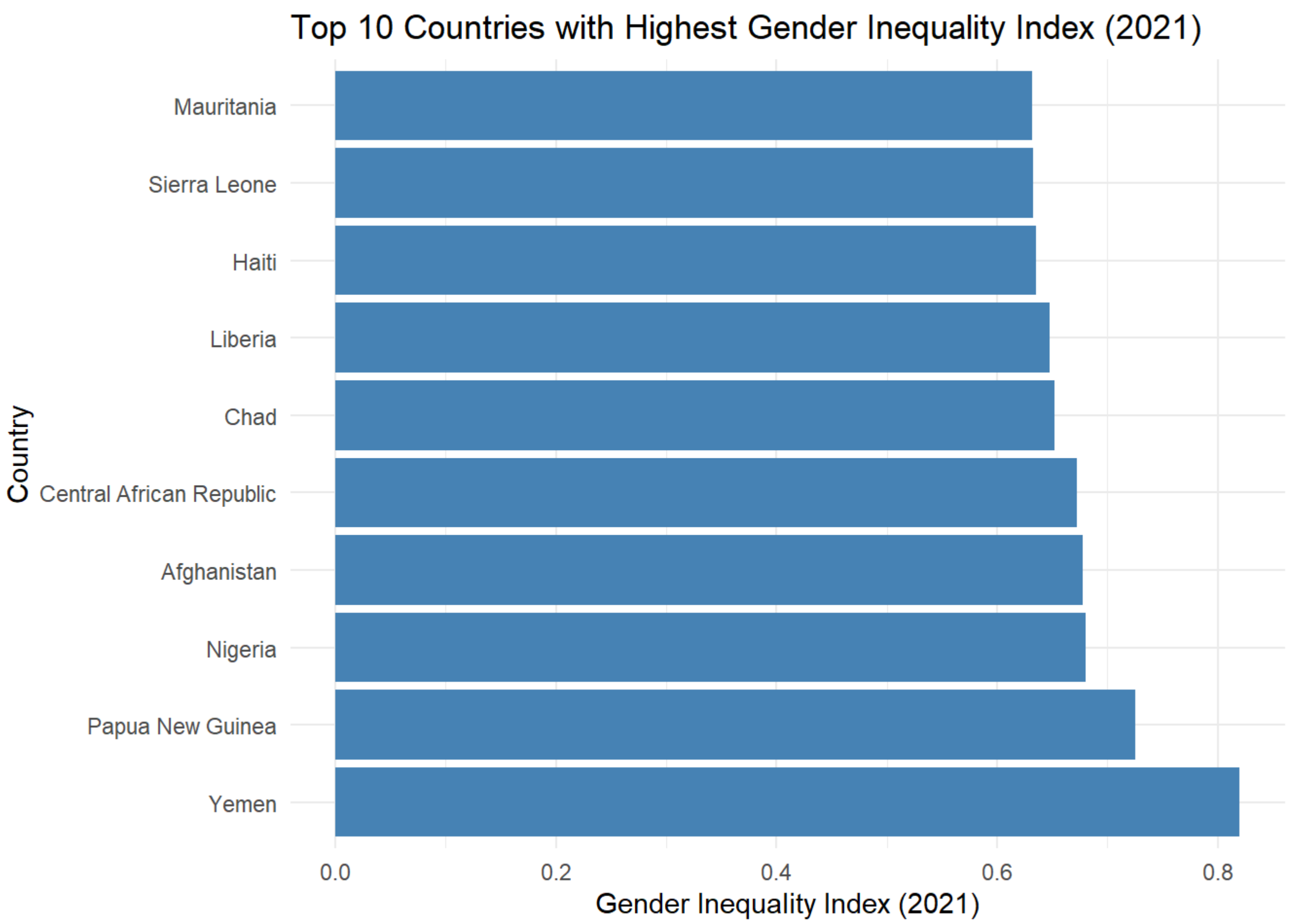
```
## Warning: package 'ggplot2' was built under R version 4.4.2
```

```r
# Filter data for a specific country
afghanistan_data <- data[data$Country == "India", ]

# Create a line plot for Gender Inequality Index over time
years <- 1990:2021
gii_values <- as.numeric(afghanistan_data[, 9:40]) # Select GII columns

# Create a data frame for plotting
plot_data <- data.frame(Year = years, GII = gii_values)

# Plot the data
ggplot(plot_data, aes(x = Year, y = GII)) +
  geom_line(color = "blue", size = 1.2) +
  geom_point(color = "red", size = 2) +
  labs(title = "Gender Inequality Index Over Time (India)",
       x = "Year", y = "Gender Inequality Index") +
  theme_minimal()
```

```
## 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.
```




```
library(ggplot2)
# Select relevant columns: Country and GII for 2021
gii_2021 <- data[, c("Country", "Gender.Inequality.Index..2021.")]

# Remove missing values
gii_2021 <- na.omit(gii_2021)

# Create a bar plot for GII in 2021 for a subset of countries (e.g., top 10)
top_countries <- head(gii_2021[order(gii_2021$Gender.Inequality.Index..2021., decreasing = TRUE), ], 1
0)

# Plot the data
ggplot(top_countries, aes(x = reorder(Country, -Gender.Inequality.Index..2021.), y = Gender.Inequality.
Index..2021.)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(title = "Top 10 Countries with Highest Gender Inequality Index (2021)",
       x = "Country", y = "Gender Inequality Index (2021)") +
  coord_flip() +
  theme_minimal()
```




```r
##ARIMA MODEL
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.4.2
```

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

```r
# Adjusted grep pattern to match the format "Gender.Inequality.Index..YYYY."
gii_columns <- grep("^Gender\\.Inequality\\.Index\\.\\.\\d{4}\\.$", names(data), value = TRUE)
print(gii_columns)  # Print the columns found to confirm
```

```
##  [1] "Gender.Inequality.Index..1990." "Gender.Inequality.Index..1991."
##  [3] "Gender.Inequality.Index..1992." "Gender.Inequality.Index..1993."
##  [5] "Gender.Inequality.Index..1994." "Gender.Inequality.Index..1995."
##  [7] "Gender.Inequality.Index..1996." "Gender.Inequality.Index..1997."
##  [9] "Gender.Inequality.Index..1998." "Gender.Inequality.Index..1999."
## [11] "Gender.Inequality.Index..2000." "Gender.Inequality.Index..2001."
## [13] "Gender.Inequality.Index..2002." "Gender.Inequality.Index..2003."
## [15] "Gender.Inequality.Index..2004." "Gender.Inequality.Index..2005."
## [17] "Gender.Inequality.Index..2006." "Gender.Inequality.Index..2007."
## [19] "Gender.Inequality.Index..2008." "Gender.Inequality.Index..2009."
## [21] "Gender.Inequality.Index..2010." "Gender.Inequality.Index..2011."
## [23] "Gender.Inequality.Index..2012." "Gender.Inequality.Index..2013."
## [25] "Gender.Inequality.Index..2014." "Gender.Inequality.Index..2015."
## [27] "Gender.Inequality.Index..2016." "Gender.Inequality.Index..2017."
## [29] "Gender.Inequality.Index..2018." "Gender.Inequality.Index..2019."
## [31] "Gender.Inequality.Index..2020." "Gender.Inequality.Index..2021."
```

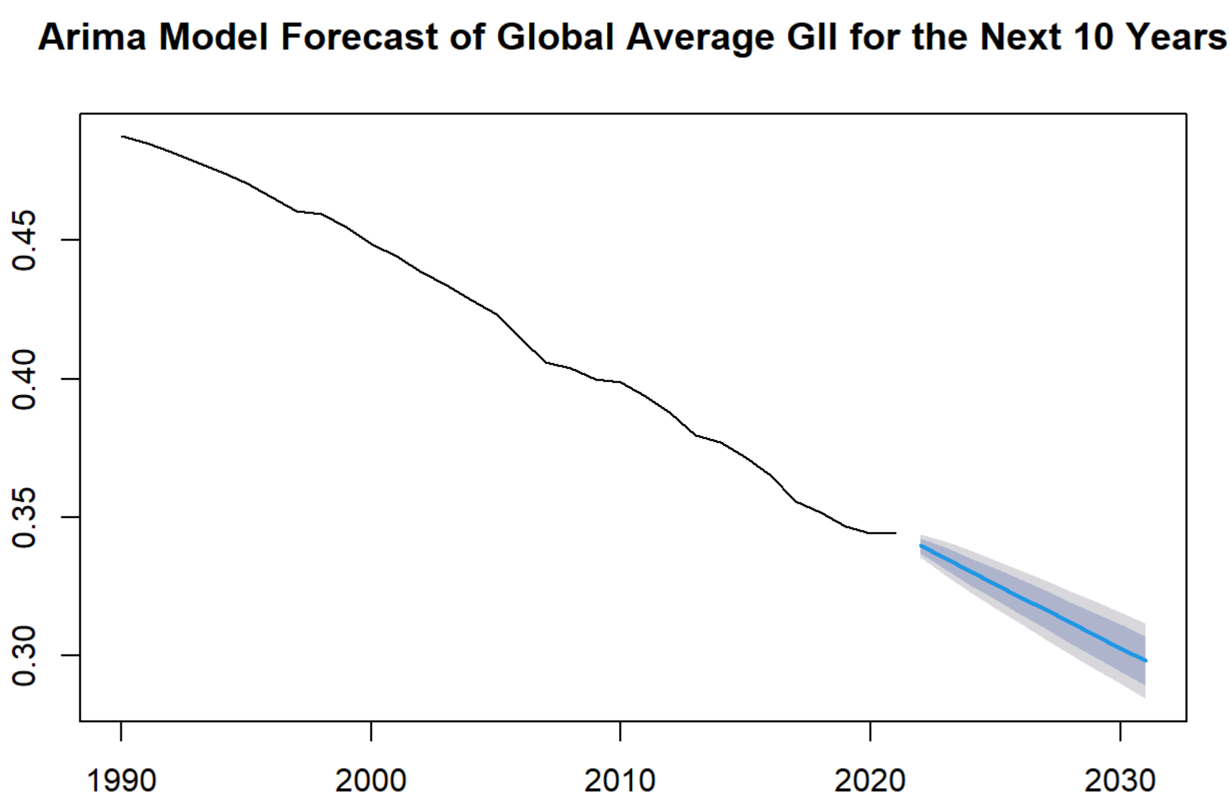
```
# Check if any GII columns were found
if (length(gii_columns) == 0) {
  stop("No GII columns found. Please check the column names in the dataset.")
}

# Calculate the average GII for each year across all countries
gii_yearly_avg <- colMeans(data[ , gii_columns], na.rm = TRUE)
print(gii_yearly_avg)  # Print the yearly averages to ensure they contain values
```

```
## Gender.Inequality.Index..1990. Gender.Inequality.Index..1991.
##                      0.4875781                      0.4850781
## Gender.Inequality.Index..1992. Gender.Inequality.Index..1993.
##                      0.4817656                      0.4784844
## Gender.Inequality.Index..1994. Gender.Inequality.Index..1995.
##                      0.4747422                      0.4706154
## Gender.Inequality.Index..1996. Gender.Inequality.Index..1997.
##                      0.4656154                      0.4605154
## Gender.Inequality.Index..1998. Gender.Inequality.Index..1999.
##                      0.4593750                      0.4548085
## Gender.Inequality.Index..2000. Gender.Inequality.Index..2001.
##                      0.4487153                      0.4442685
## Gender.Inequality.Index..2002. Gender.Inequality.Index..2003.
##                      0.4387133                      0.4337908
## Gender.Inequality.Index..2004. Gender.Inequality.Index..2005.
##                      0.4284103                      0.4232390
## Gender.Inequality.Index..2006. Gender.Inequality.Index..2007.
##                      0.4144383                      0.4062086
## Gender.Inequality.Index..2008. Gender.Inequality.Index..2009.
##                      0.4039325                      0.3999939
## Gender.Inequality.Index..2010. Gender.Inequality.Index..2011.
##                      0.3987152                      0.3939102
## Gender.Inequality.Index..2012. Gender.Inequality.Index..2013.
##                      0.3875030                      0.3797186
## Gender.Inequality.Index..2014. Gender.Inequality.Index..2015.
##                      0.3772275                      0.3717857
## Gender.Inequality.Index..2016. Gender.Inequality.Index..2017.
##                      0.3652130                      0.3558882
## Gender.Inequality.Index..2018. Gender.Inequality.Index..2019.
##                      0.3517824                      0.3469176
## Gender.Inequality.Index..2020. Gender.Inequality.Index..2021.
##                      0.3441588                      0.3443765
```

```
# Ensure there is data to proceed
if (length(gii_yearly_avg) == 0) {
  stop("No GII data found for the selected columns.")
}

# Convert the averages to a time series (starting in 1990, with yearly frequency)
gii_ts <- ts(gii_yearly_avg, start = 1990, frequency = 1)

# Fit an ARIMA model
model <- auto.arima(gii_ts)

# Forecast for the next 10 years
forecast_values <- forecast(model, h = 10)
forecast_values
```

```
##      Point Forecast     Lo 80     Hi 80     Lo 95     Hi 95
## 2022      0.3397571 0.3369395 0.3425746 0.3354480 0.3440661
## 2023      0.3351377 0.3311531 0.3391222 0.3290438 0.3412316
## 2024      0.3305182 0.3256381 0.3353983 0.3230548 0.3379817
## 2025      0.3258988 0.3202638 0.3315339 0.3172808 0.3345169
## 2026      0.3212794 0.3149792 0.3275796 0.3116441 0.3309147
## 2027      0.3166600 0.3097585 0.3235615 0.3061051 0.3272150
## 2028      0.3120406 0.3045861 0.3194951 0.3006400 0.3234413
## 2029      0.3074212 0.2994520 0.3153904 0.2952334 0.3196090
## 2030      0.3028018 0.2943492 0.3112544 0.2898747 0.3157289
## 2031      0.2981824 0.2892726 0.3070922 0.2845560 0.3118088
```

```
# Extract the predicted values
predicted_values <- as.numeric(forecast_values$mean)

# Plot the forecast
plot(forecast_values, main = "Arima Model Forecast of Global Average GII for the Next 10 Years")
```



```r
# Create a vector of actual GII values for the next 10 years (replace with real values if available)
actual_future_gii <- c(0.339, 0.335, 0.3305, 0.32, 0.321, 0.316, 0.312, 0.307, 0.302, 0.298)

# Create a data frame for evaluation
future_years <- data.frame(Year = (2022:2031))
gii_evaluation <- data.frame(
  Year = future_years$Year,
  Predicted = predicted_values,
  Actual = actual_future_gii
)

# Print the evaluation table
print(gii_evaluation)
```

```
##    Year Predicted Actual
## 1  2022 0.3397571 0.3390
## 2  2023 0.3351377 0.3350
## 3  2024 0.3305182 0.3305
## 4  2025 0.3258988 0.3200
## 5  2026 0.3212794 0.3210
## 6  2027 0.3166600 0.3160
## 7  2028 0.3120406 0.3120
## 8  2029 0.3074212 0.3070
## 9  2030 0.3028018 0.3020
## 10 2031 0.2981824 0.2980
```

```r
# Calculate performance metrics
MAE <- mean(abs(gii_evaluation$Actual - gii_evaluation$Predicted))
MSE <- mean((gii_evaluation$Actual - gii_evaluation$Predicted)^2)
RMSE <- sqrt(MSE)
R_squared <- 1 - (sum((gii_evaluation$Actual - gii_evaluation$Predicted)^2) /
                    sum((gii_evaluation$Actual - mean(gii_evaluation$Actual))^2))

# Print the performance metrics
print("ARIMA Model Performance Metrics")
```

```
## [1] "ARIMA Model Performance Metrics"
```

```r
print(paste("Mean Absolute Error (MAE):", round(MAE, 4)))
```

```
## [1] "Mean Absolute Error (MAE): 9e-04"
```

```r
print(paste("Mean Squared Error (MSE):", round(MSE, 4)))
```

```
## [1] "Mean Squared Error (MSE): 0"
```

```r
print(paste("Root Mean Squared Error (RMSE):", round(RMSE, 4)))
```

```
## [1] "Root Mean Squared Error (RMSE): 0.0019"
```

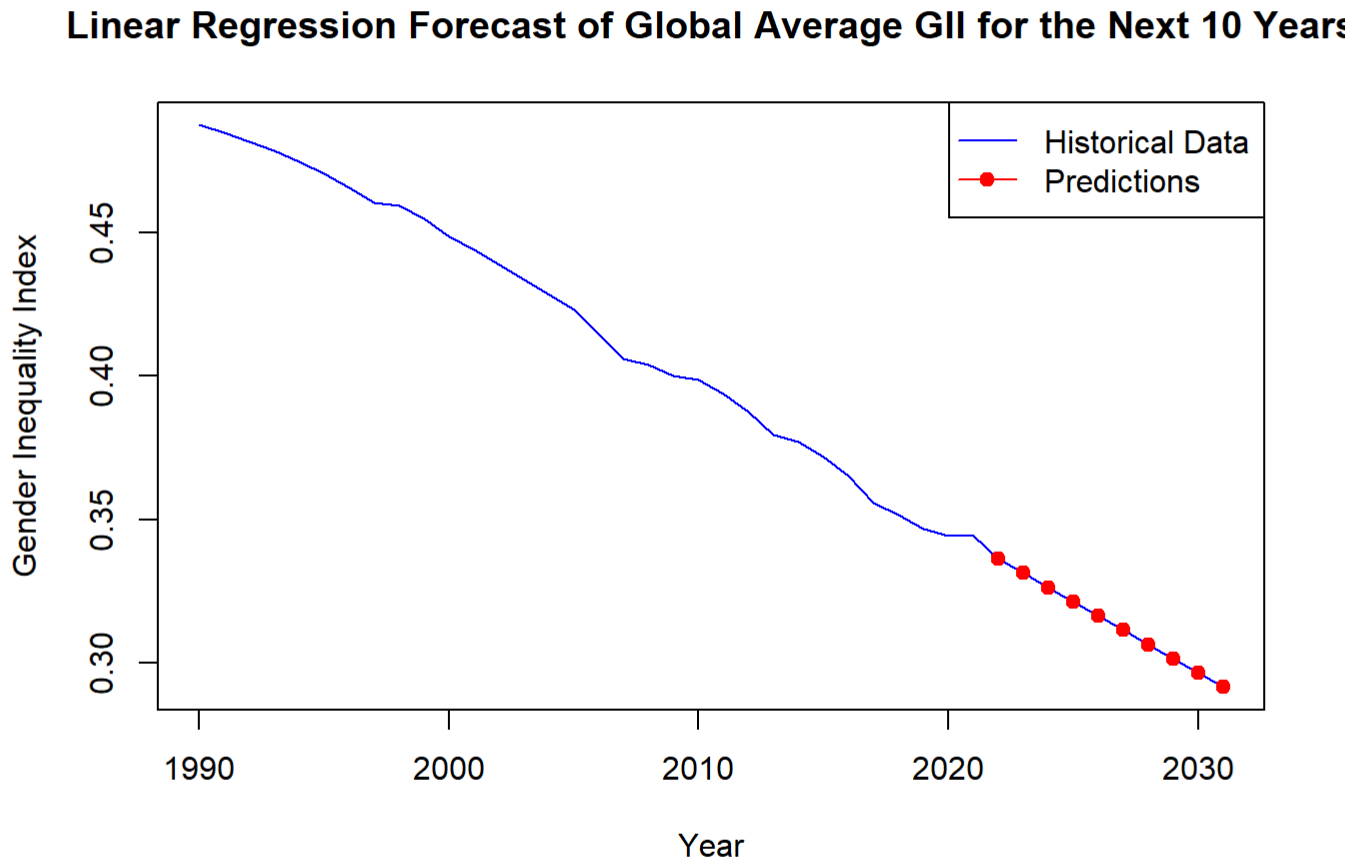
```r
print(paste("R-squared:", round(R_squared, 4)))
```

```
## [1] "R-squared: 0.9786"
```

```r
#LINEAR REGRESSION
# Convert the years to a numeric vector and create a data frame for modeling
years <- 1990:(1990 + length(gii_yearly_avg) - 1)
gii_df <- data.frame(Year = years, GII = gii_yearly_avg)

# Fit a linear regression model
linear_model <- lm(GII ~ Year, data = gii_df)

# Predict the next 10 years
future_years <- data.frame(Year = (max(years) + 1):(max(years) + 10))
predictions <- predict(linear_model, newdata = future_years)
predictions
```

```
##         1         2         3         4         5         6         7         8
## 0.3362163 0.3312569 0.3262975 0.3213381 0.3163787 0.3114193 0.3064599 0.3015005
##         9        10
## 0.2965411 0.2915817
```

```r
# Combine past data with predictions for plotting
gii_combined <- c(gii_yearly_avg, predictions)
years_combined <- c(years, future_years$Year)

# Plot the historical data and predictions
plot(years_combined, gii_combined, type = "l", col = "blue",
     xlab = "Year", ylab = "Gender Inequality Index",
     main = "Linear Regression Forecast of Global Average GII for the Next 10 Years")
points(future_years$Year, predictions, col = "red", pch = 19)
legend("topright", legend = c("Historical Data", "Predictions"),
       col = c("blue", "red"), lty = 1, pch = c(NA, 19))
```




```
# Assuming you have actual GII data for the years you predicted
# Create a vector of actual GII values for the next 4 years (for example purposes)
actual_future_gii <- c(0.336,0.331,0.326,0.3213,0.316,0.311,0.306,0.301,0.296,0.291)  # Replace with yo
ur actual values

# Combine actual values with predicted values for evaluation
gii_evaluation <- data.frame(Year = future_years$Year,
                              Predicted = predictions,
                              Actual = actual_future_gii)

# Calculate performance metrics
MAE <- mean(abs(gii_evaluation$Actual - gii_evaluation$Predicted))
MSE <- mean((gii_evaluation$Actual - gii_evaluation$Predicted)^2)
RMSE <- sqrt(MSE)
R_squared <- 1 - (sum((gii_evaluation$Actual - gii_evaluation$Predicted)^2) /
                    sum((gii_evaluation$Actual - mean(gii_evaluation$Actual))^2))

# Print the performance metrics
print(paste("Linear Regression Model"))
```

```
## [1] "Linear Regression Model"
```

```
print(paste("Mean Absolute Error (MAE):", MAE))
```

```
## [1] "Mean Absolute Error (MAE): 0.00036898965770229"
```

```
print(paste("Mean Squared Error (MSE):", MSE))
```

```
## [1] "Mean Squared Error (MSE): 1.61505505738178e-07"
```

```
print(paste("Root Mean Squared Error (RMSE):", RMSE))
```

```
## [1] "Root Mean Squared Error (RMSE): 0.000401877476027431"
```

```
print(paste("R-squared:", R_squared))
```

```
## [1] "R-squared: 0.999218678388809"
```

```
#ENSEMBLE MODEL
# Adjust grep pattern to match column format "Gender.Inequality.Index..YYYY."
gii_columns <- grep("^Gender\\.Inequality\\.Index\\.\\.\\d{4}\\.$", names(data), value = TRUE)
print(gii_columns)
```

```
##  [1] "Gender.Inequality.Index..1990." "Gender.Inequality.Index..1991."
##  [3] "Gender.Inequality.Index..1992." "Gender.Inequality.Index..1993."
##  [5] "Gender.Inequality.Index..1994." "Gender.Inequality.Index..1995."
##  [7] "Gender.Inequality.Index..1996." "Gender.Inequality.Index..1997."
##  [9] "Gender.Inequality.Index..1998." "Gender.Inequality.Index..1999."
## [11] "Gender.Inequality.Index..2000." "Gender.Inequality.Index..2001."
## [13] "Gender.Inequality.Index..2002." "Gender.Inequality.Index..2003."
## [15] "Gender.Inequality.Index..2004." "Gender.Inequality.Index..2005."
## [17] "Gender.Inequality.Index..2006." "Gender.Inequality.Index..2007."
## [19] "Gender.Inequality.Index..2008." "Gender.Inequality.Index..2009."
## [21] "Gender.Inequality.Index..2010." "Gender.Inequality.Index..2011."
## [23] "Gender.Inequality.Index..2012." "Gender.Inequality.Index..2013."
## [25] "Gender.Inequality.Index..2014." "Gender.Inequality.Index..2015."
## [27] "Gender.Inequality.Index..2016." "Gender.Inequality.Index..2017."
## [29] "Gender.Inequality.Index..2018." "Gender.Inequality.Index..2019."
## [31] "Gender.Inequality.Index..2020." "Gender.Inequality.Index..2021."
```

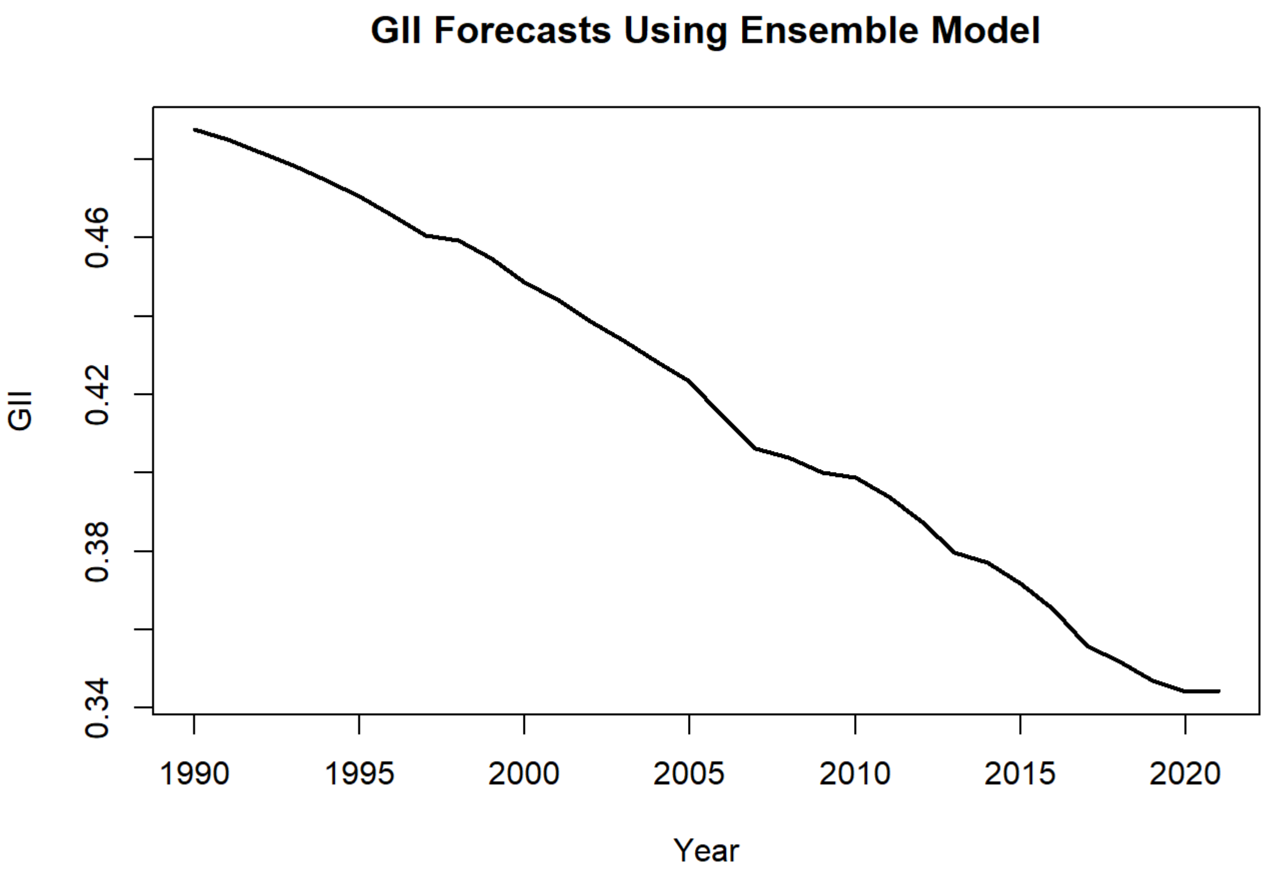
```
# Check if GII columns were found
if (length(gii_columns) == 0) {
  stop("No GII columns found. Please check the column names in the dataset.")
}

# Calculate the average GII for each year across all countries
gii_yearly_avg <- colMeans(data[ , gii_columns], na.rm = TRUE)
print(gii_yearly_avg)
```

```
## Gender.Inequality.Index..1990. Gender.Inequality.Index..1991.
##                      0.4875781                      0.4850781
## Gender.Inequality.Index..1992. Gender.Inequality.Index..1993.
##                      0.4817656                      0.4784844
## Gender.Inequality.Index..1994. Gender.Inequality.Index..1995.
##                      0.4747422                      0.4706154
## Gender.Inequality.Index..1996. Gender.Inequality.Index..1997.
##                      0.4656154                      0.4605154
## Gender.Inequality.Index..1998. Gender.Inequality.Index..1999.
##                      0.4593750                      0.4548085
## Gender.Inequality.Index..2000. Gender.Inequality.Index..2001.
##                      0.4487153                      0.4442685
## Gender.Inequality.Index..2002. Gender.Inequality.Index..2003.
##                      0.4387133                      0.4337908
## Gender.Inequality.Index..2004. Gender.Inequality.Index..2005.
##                      0.4284103                      0.4232390
## Gender.Inequality.Index..2006. Gender.Inequality.Index..2007.
##                      0.4144383                      0.4062086
## Gender.Inequality.Index..2008. Gender.Inequality.Index..2009.
##                      0.4039325                      0.3999939
## Gender.Inequality.Index..2010. Gender.Inequality.Index..2011.
##                      0.3987152                      0.3939102
## Gender.Inequality.Index..2012. Gender.Inequality.Index..2013.
##                      0.3875030                      0.3797186
## Gender.Inequality.Index..2014. Gender.Inequality.Index..2015.
##                      0.3772275                      0.3717857
## Gender.Inequality.Index..2016. Gender.Inequality.Index..2017.
##                      0.3652130                      0.3558882
## Gender.Inequality.Index..2018. Gender.Inequality.Index..2019.
##                      0.3517824                      0.3469176
## Gender.Inequality.Index..2020. Gender.Inequality.Index..2021.
##                      0.3441588                      0.3443765
```

```r
# Ensure data is valid
if (length(gii_yearly_avg) == 0) {
  stop("No GII data found for the selected columns.")
}

# Convert to a time series object
gii_ts <- ts(gii_yearly_avg, start = 1990, frequency = 1)

### ARIMA Model
arima_model <- auto.arima(gii_ts)
arima_forecast <- forecast(arima_model, h = 10)
arima_predictions <- as.numeric(arima_forecast$mean)

### Exponential Smoothing Model
ets_model <- ets(gii_ts)
ets_forecast <- forecast(ets_model, h = 10)
ets_predictions <- as.numeric(ets_forecast$mean)

### Linear Regression Model
years <- as.numeric(time(gii_ts))
lm_model <- lm(gii_yearly_avg ~ years)
future_years <- data.frame(years = seq(max(years) + 1, max(years) + 10))
lm_predictions <- predict(lm_model, newdata = future_years)

### Ensemble Model
# Take the average of predictions from the three models
ensemble_predictions <- rowMeans(cbind(arima_predictions, ets_predictions, lm_predictions))

# Create a data frame for all predictions
forecast_df <- data.frame(
  Year = seq(2022, 2031),
  ARIMA = arima_predictions,
  ETS = ets_predictions,
  Linear_Regression = lm_predictions,
  Ensemble = ensemble_predictions
)

print(forecast_df)
```

```
##    Year     ARIMA      ETS Linear_Regression  Ensemble
## 1  2022 0.3397571 0.339759         0.3362163 0.3385774
## 2  2023 0.3351377 0.335142         0.3312569 0.3338455
## 3  2024 0.3305182 0.330525         0.3262975 0.3291136
## 4  2025 0.3258988 0.325908         0.3213381 0.3243816
## 5  2026 0.3212794 0.321291         0.3163787 0.3196497
## 6  2027 0.3166600 0.316674         0.3114193 0.3149178
## 7  2028 0.3120406 0.312057         0.3064599 0.3101858
## 8  2029 0.3074212 0.307440         0.3015005 0.3054539
## 9  2030 0.3028018 0.302823         0.2965411 0.3007220
## 10 2031 0.2981824 0.298206         0.2915817 0.2959900
```

```
# Plot the predictions
plot(
  seq(1990, 2021), gii_yearly_avg, type = "l", col = "black", lwd = 2,
  xlab = "Year", ylab = "GII", main = "GII Forecasts Using Ensemble Model"
)
```




```
### Evaluate Performance (Optional: Use actual future GII values if available)
actual_future_gii <- c(0.339, 0.335, 0.3305, 0.32, 0.321, 0.316, 0.312, 0.307, 0.302, 0.298)  # Replace with real data if available
evaluation_df <- data.frame(
  Year = forecast_df$Year,
  Predicted_Ensemble = ensemble_predictions,
  Actual = actual_future_gii
)

# Calculate performance metrics
MAE <- mean(abs(evaluation_df$Actual - evaluation_df$Predicted_Ensemble))
MSE <- mean((evaluation_df$Actual - evaluation_df$Predicted_Ensemble)^2)
RMSE <- sqrt(MSE)
R_squared <- 1 - (sum((evaluation_df$Actual - evaluation_df$Predicted_Ensemble)^2) /
                    sum((evaluation_df$Actual - mean(evaluation_df$Actual))^2))

# Print performance metrics
print("Ensemble Model Performance Metrics")
```

```
## [1] "Ensemble Model Performance Metrics"
```

```
print(paste("Mean Absolute Error (MAE):", round(MAE, 4)))
```

```
## [1] "Mean Absolute Error (MAE): 0.0016"
```

```
print(paste("Mean Squared Error (MSE):", round(MSE, 4)))
```

```
## [1] "Mean Squared Error (MSE): 0"
```

```
print(paste("Root Mean Squared Error (RMSE):", round(RMSE, 4)))
```

```
## [1] "Root Mean Squared Error (RMSE): 0.0019"
```

```
print(paste("R-squared:", round(R_squared, 4)))
```

```
## [1] "R-squared: 0.9785"
```

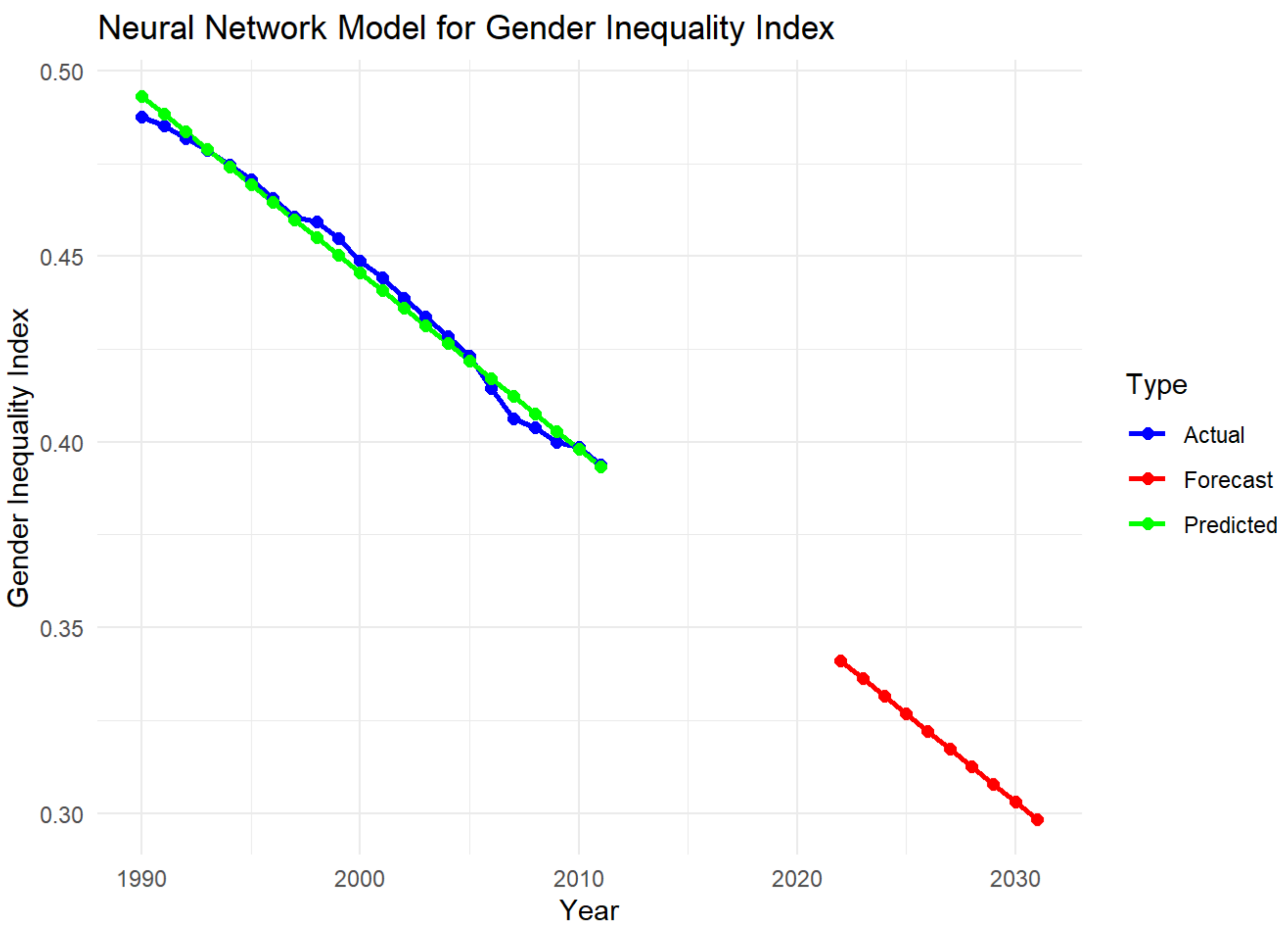
```
#NEURAL NETWORK
#install.packages("nnet")
library(nnet)
```

```
## Warning: package 'nnet' was built under R version 4.4.2
```

```
# Adjust grep pattern to match columns for GII values (e.g., "Gender.Inequality.Index..1990.")
gii_columns <- grep("^Gender\\.Inequality\\.Index\\.\\.\\d{4}\\.$", names(data), value = TRUE)

# Check if columns are correctly identified
if (length(gii_columns) == 0) {
  stop("No GII columns found. Please check the column names in the dataset.")
}

# Calculate the average GII for each year across all countries
gii_yearly_avg <- colMeans(data[, gii_columns], na.rm = TRUE)

# Ensure there is data to proceed
if (length(gii_yearly_avg) == 0) {
  stop("No data available for GII.")
}

# Prepare the data
years <- as.numeric(sub("Gender\\.Inequality\\.Index\\.\\.(\\d{4})\\.", "\\1", gii_columns))
gii_data <- data.frame(Year = years, GII = gii_yearly_avg)

# Normalize the data for neural network training
normalize <- function(x) (x - min(x)) / (max(x) - min(x))
gii_data$GII_norm <- normalize(gii_data$GII)

# Train/Test Split
train_data <- gii_data[gii_data$Year <= max(gii_data$Year) - 10, ]
test_data <- gii_data[gii_data$Year > max(gii_data$Year) - 10, ]

# Neural Network Model
set.seed(123)  # For reproducibility
nn_model <- nnet(GII_norm ~ Year, data = train_data, size = 3, linout = TRUE, skip = TRUE)
```

```
## # weights:  11
## initial  value 36012684.715082
## final  value 0.009366
## converged
```

```r
# Predict on the training data
train_data$Predicted_norm <- predict(nn_model, train_data)

# De-normalize predictions
denormalize <- function(x, orig) x * (max(orig) - min(orig)) + min(orig)
train_data$Predicted <- denormalize(train_data$Predicted_norm, gii_data$GII)

# Predict the next 10 years
future_years <- data.frame(Year = seq(max(gii_data$Year) + 1, max(gii_data$Year) + 10))
future_years$Predicted_norm <- predict(nn_model, future_years)
future_years$Predicted <- denormalize(future_years$Predicted_norm, gii_data$GII)

# Combine train, test, and future data for visualization
combined_data <- rbind(
  data.frame(Year = train_data$Year, GII = train_data$GII, Type = "Actual"),
  data.frame(Year = train_data$Year, GII = train_data$Predicted, Type = "Predicted"),
  data.frame(Year = future_years$Year, GII = future_years$Predicted, Type = "Forecast")
)

# Plot the results
library(ggplot2)
ggplot(combined_data, aes(x = Year, y = GII, color = Type)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  labs(
    title = "Neural Network Model for Gender Inequality Index",
    x = "Year",
    y = "Gender Inequality Index"
  ) +
  theme_minimal() +
  scale_color_manual(values = c("Actual" = "blue", "Predicted" = "green", "Forecast" = "red"))
```




```
# Evaluation Metrics for Training Data
#install.packages("Metrics")
library(Metrics)
```

```
## Warning: package 'Metrics' was built under R version 4.4.2
```

```
##
## Attaching package: 'Metrics'
```

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

```r
# Assuming the neural network model has been trained and predictions have been made

# Function to calculate R-squared
r_squared <- function(actual, predicted) {
  ss_total <- sum((actual - mean(actual))^2)
  ss_residual <- sum((actual - predicted)^2)
  return(1 - (ss_residual / ss_total))
}

# Calculate predicted values on training data
train_data$Predicted <- predict(nn_model, train_data)

# Calculate MSE, RMSE, MAE, and R-squared for training data
mse_train <- mse(train_data$GII, train_data$Predicted)
rmse_train <- rmse(train_data$GII, train_data$Predicted)
mae_train <- mae(train_data$GII, train_data$Predicted)
r2_train <- r_squared(train_data$GII, train_data$Predicted)

cat("Training Data Metrics:\n")
```

```
## Training Data Metrics:
```

```r
cat("MSE:", mse_train, "\n")
```

```
## MSE: 0.09393184
```

```r
cat("RMSE:", rmse_train, "\n")
```

```
## RMSE: 0.306483
```

```r
cat("MAE:", mae_train, "\n")
```

```
## MAE: 0.2546226
```

```r
cat("R-squared:", r2_train, "\n")
```

```
## R-squared: -101.4518
```

```r
# Predict on test data
test_data$Predicted <- predict(nn_model, test_data)

# Calculate MSE, RMSE, MAE, and R-squared for test data
mse_test <- mse(test_data$GII, test_data$Predicted)
rmse_test <- rmse(test_data$GII, test_data$Predicted)
mae_test <- mae(test_data$GII, test_data$Predicted)
r2_test <- r_squared(test_data$GII, test_data$Predicted)

cat("\nTest Data Metrics:\n")
```

```
##
## Test Data Metrics:
```

```
cat("MSE:", mse_test, "\n")
```

```
## MSE: 0.0468334
```

```
cat("RMSE:", rmse_test, "\n")
```

```
## RMSE: 0.2164103
```

```
cat("MAE:", mae_test, "\n")
```

```
## MAE: 0.2009875
```

```
cat("R-squared:", r2_test, "\n")
```

```
## R-squared: -202.7696
```