# Deforestation-to-biodiversity

Ulas Ayyilmaz and Ishika

```r
#change between 2000-2020 #https://www.globalforestwatch.org/dashboards/global/?category=fore
forest_data <- read_csv("net_tree_change.csv")
head(forest_data)
```

```
# A tibble: 6 x 8
  iso   stable                   loss    gain  disturb net   change gfw_area__ha
  <chr> <chr>                    <chr>   <chr> <chr>   <chr>  <dbl>        <dbl>
1 ABW   113.038689599999998      6.7644~ 19.2~ 0.6764~ 12.4~  10.4         18194.
2 AFG   368081.866700000013225   16604.~ 1074~ 1146.2~ -586~  -1.52    64385933.
3 AGO   44546560.189999997615814 341261~ 1224~ 162392~ -218~  -4.41   124742581.
4 AIA   890.152399199999991      51.507~ 69.8~ 9.7319~ 18.3~   1.93         8330.
5 ALA   76419.028500000000349    7426.9~ 2582~ 12417.~ -484~  -5.03      150631.
6 ALB   814631.915999999968335   40701.~ 1647~ 41219.~ -242~  -2.70     2873543.
```

```r
top_5_loss <- forest_data |>
  arrange(desc(as.numeric(loss))) |>
  slice_head(n = 20)
top_5_gain <- forest_data |>
  arrange(desc(as.numeric(gain))) |>
  slice_head(n = 20)

top_5_net_desc <- forest_data |>
  arrange(desc(as.numeric(net))) |>
  slice_head(n = 20)

top_5_net_asc <- forest_data |>
  arrange(as.numeric(net)) |>
  slice_head(n = 20)

top_5_net_desc # poland, ukraine, uruguay, ireland, bangladesh
```

```
# A tibble: 20 x 8
   iso   stable                      loss  gain  disturb net   change gfw_area__ha
   <chr> <chr>                       <chr> <chr> <chr>   <chr>  <dbl>        <dbl>
 1 CHN   202215007.599999994039536 4544~ 6689~ 171769~ 2144~  0.958   938225140.
 2 IND   59241391.750000000000000  1007~ 1881~ 617156~ 8740~  1.32    315280003.
 3 URY   659122.222900000051595    1127~ 6544~ 230107~ 5417~ 54.1      17746563.
 4 BLR   7407088.671000000089407   4421~ 9636~ 683220~ 5214~  6.11     20706981.
 5 UKR   9065233.101999999955297   4998~ 9263~ 895296~ 4265~  4.08     60136264.
 6 POL   9114498.453999999910593   4853~ 8920~ 975192~ 4066~  3.85     31240006.
 7 SSD   25841911.940000001341105  8019~ 1082~ 96714.~ 2809~  1.05     62810211.
 8 BGD   3313032.510999999940395   1154~ 3178~ 612964~ 2024~  5.01     13938445.
 9 IRL   388037.844199999992270    9463~ 2842~ 77309.~ 1896~ 33.9       7036677.
10 SDN   3715927.944000000134110   8530~ 2165~ 2345.3~ 1312~  3.45    187212843.
11 LTU   1722243.843000000109896   1166~ 2415~ 315129~ 1248~  5.80      6501684.
12 ROU   7427728.728000000119209   1057~ 2238~ 534902~ 1181~  1.46     23833860.
13 NPL   6718407.851999999955297   7372~ 1740~ 55029.~ 1002~  1.46     14766264.
14 PAK   2057786.046000000089407   3869~ 1334~ 7457.5~ 9478~  4.51     87417714.
15 THA   18845385.170000001788139  1747~ 1842~ 405750~ 9441~  0.383    51405456.
16 TUR   12756770.300000000745058  3850~ 4737~ 510036~ 8876~  0.650    78070436.
17 HUN   1677658.310000000055879   1226~ 2073~ 249845~ 8463~  4.13      9305287.
18 GBR   2421686.652999999932945   3915~ 4731~ 184814~ 8156~  2.72     24548584.
19 GEO   3238385.424999999813735   1795~ 9353~ 7227.5~ 7558~  2.32      6984461.
20 PRK   6061637.154000000096858   2135~ 2884~ 394941~ 7497~  1.12     12275513.
```

```r
top_5_net_asc#tanzania, Mozambique, indonesia, DCcongo, paraguay
```

```
# A tibble: 20 x 8
   iso   stable                      loss  gain  disturb net   change gfw_area__ha
   <chr> <chr>                       <chr> <chr> <chr>   <chr>  <dbl>        <dbl>
 1 BRA   413722809.300000011920929 3614~ 8062~ 234216~ -280~  -5.93   850036547.
 2 CAN   257102961.500000000000000 2516~ 1696~ 148636~ -820~  -2.76   995519376.
 3 COD   146862160.800000011920929 7593~ 1591~ 145662~ -600~  -3.55   232913006.
 4 PRY   14232500.500000000000000  5810~ 6424~ 883902~ -516~ -24.7     39958592.
 5 MOZ   35317855.280000001192093  4897~ 5831~ 202499~ -431~ -10.2     78729703.
 6 IDN   124187188.599999994039536 9000~ 4882~ 240569~ -411~  -2.62   189024479.
 7 TZA   27968218.850000001490116  4374~ 5572~ 143782~ -381~ -11.3     94057605.
 8 ARG   27610827.399999998509884  4664~ 1107~ 174446~ -355~ -10.5    278009661.
 9 USA   237725811.800000011920929 1747~ 1398~ 283653~ -348~  -1.23   947301497.
10 BOL   52561114.219999998807907  3941~ 6172~ 270692~ -332~  -5.61   108339299.
11 ZMB   34159461.780000001192093  3640~ 7681~ 151400~ -287~  -7.31    75049045.
12 KHM   6195015.302000000141561   2787~ 1473~ 133709~ -264~ -25.6     18135962.
13 AGO   44546560.189999997615814  3412~ 1224~ 162392~ -218~  -4.41   124742581.
```

```
14 COL   73177488.349999994039536  2825~ 1085~ 328654~ -174~  -2.20   113675882.
15 CIV   16844645.399999998509884  2530~ 8816~ 437764~ -164~  -6.94    32165807.
16 MMR   35420448.899999998509884  2227~ 6366~ 839121~ -159~  -3.46    66929741.
17 GIN   14569792.289999999105930  1862~ 2726~ 262048~ -158~  -8.34    24481617.
18 LAO   13290214.820000000298023  1573~ 1017~ 568687~ -147~  -7.16    23000163.
19 NGA   20636992.129999998956919  2396~ 9278~ 118123~ -146~  -6.07    90841014.
20 VEN   51803026.079999998211861  1860~ 4913~ 133297~ -136~  -2.49    91249272.
```

species of interest (check exist for each country)

Most net negative: PRY, COD, MOZ, IDN, TZA

most positive: URY, UKR, POL,IRL, BGD

All columns of a RGBIF EOD dataset #"gbifID", "datasetKey", "occurrenceID", "kingdom", "phylum", "class", "order", "family", #"genus", "species", "infraspecificEpithet", "taxonRank", "scientificName", #"verbatimScientificName", "verbatimScientificNameAuthorship", "countryCode", "locality", #"stateProvince", "occurrenceStatus", "individualCount", "publishingOrgKey", #"decimalLatitude", "decimalLongitude", "coordinateUncertaintyInMeters", #"coordinatePrecision", "elevation", "elevationAccuracy", "depth", "depthAccuracy", #"eventDate", "day", "month", "year", "taxonKey", "speciesKey", "basisOfRecord", #"institutionCode", "collectionCode", "catalogNumber", "recordNumber", "identifiedBy", #"dateIdentified", "license", "rightsHolder", "recordedBy", "typeStatus", #"establishmentMeans", "lastInterpreted", "mediaType", "issue")

Ok, brainstorm time. Deforestation means removel of forests over time due to mostly due to human interference. Deforestation affects many things that contribute to human's well-being indirectly in a negative way. A direct effect is observed on the biodiversity - specifically birds who roam freely in forests. Other

```r
# Parameters
countries <- c("RUS", "BRA", "CAN", "USA", "CHN", "IDN", "IND", "URY", "BLR", "UKR")  # Targ
years <- 2000:2020  # Year range
bird_data<-name_suggest(q = "Aves", rank = "class", curlopts = list(timeout = 60))
bird_taxon_key <- bird_data$data$key
eod_dataset_key <- "4fa7b334-ce0d-4e88-aaae-2e0c138d049e"  # EOD datasetKey
bird_taxon_key
```

```
[1] 212
```

```r
# Define the target countries and years
countries <- c("URY","BLR","UKR","POL","SSD")  # Target countries
years <- c(2000, 2010, 2020)  # Year range
```

```r
eod_dataset_key <- "4fa7b334-ce0d-4e88-aaae-2e0c138d049e"  # EOD datasetKey
species <- c("Coraciiformes","Strigiformes","Galliformes","Ciconiiformes")
```

```r
# Load necessary library
library(dplyr)  # For data manipulation

# Initialize an empty dataframe to store the combined data
combined_data <- data.frame()

# List of country names (folders inside the "data" directory)
countries <- c( "poland","ukraine", "uruguay", "ireland", "bangladesh",
               "tanzania", "mozambique", "indonesia", "dcongo", "paraguay")
countries
```

```
 [1] "poland"     "ukraine"    "uruguay"    "ireland"    "bangladesh"
 [6] "tanzania"   "mozambique" "indonesia"  "dcongo"     "paraguay"
```

```r
# Loop through each country's folder and combine all CSV files
for (country in countries) {

  # Path to the country's folder
  country_path <- file.path("data", country)

    # Get the list of CSV files in the country's folder
  csv_files <- list.files(country_path, pattern = "\\.csv$", full.names = TRUE)

    # Read each CSV file and add its content to the combined dataframe
  for (csv_file in csv_files) {
   tryCatch({
      # Read the CSV file (use read_delim for robustness)
      data <- read_delim(csv_file, delim = NULL, show_col_types = FALSE)

      # Add a column to identify the country
      data$country <- country

      # Append the data to the combined dataframe
      combined_data <- bind_rows(combined_data, data)
    }, error = function(e) {
      cat("Error reading file:", csv_file, "\n")
    })
  }
```

```r
}

# Save the combined dataframe as a CSV file in the main directory
write.csv(combined_data, "all_bird_orders_data.csv", row.names = FALSE)

cat("Combined CSV file created as 'all_bird_orders_data.csv' in the main directory.\n")
```

```
Combined CSV file created as 'all_bird_orders_data.csv' in the main directory.
```

```r
bird_orders <- read.csv("all_bird_orders_data.csv")
head(bird_orders)
```

```
     gbifID                           datasetKey
1 972752070 4fa7b334-ce0d-4e88-aaae-2e0c138d049e
2 962063576 4fa7b334-ce0d-4e88-aaae-2e0c138d049e
3 956424085 4fa7b334-ce0d-4e88-aaae-2e0c138d049e
4 956423830 4fa7b334-ce0d-4e88-aaae-2e0c138d049e
5 956334143 4fa7b334-ce0d-4e88-aaae-2e0c138d049e
6 584375475 4fa7b334-ce0d-4e88-aaae-2e0c138d049e
                        occurrenceID  kingdom   phylum class         order
1 URN:catalog:CLO:EBIRD:OBS219042257 Animalia Chordata  Aves Ciconiiformes
2 URN:catalog:CLO:EBIRD:OBS207389323 Animalia Chordata  Aves Ciconiiformes
3 URN:catalog:CLO:EBIRD:OBS200802257 Animalia Chordata  Aves Ciconiiformes
4 URN:catalog:CLO:EBIRD:OBS200798945 Animalia Chordata  Aves Ciconiiformes
5 URN:catalog:CLO:EBIRD:OBS200758554 Animalia Chordata  Aves Ciconiiformes
6  URN:catalog:CLO:EBIRD:OBS96224279 Animalia Chordata  Aves Ciconiiformes
      family   genus         species infraspecificEpithet taxonRank
1 Ciconiidae Ciconia Ciconia ciconia                   NA   SPECIES
2 Ciconiidae Ciconia   Ciconia nigra                   NA   SPECIES
3 Ciconiidae Ciconia Ciconia ciconia                   NA   SPECIES
4 Ciconiidae Ciconia Ciconia ciconia                   NA   SPECIES
5 Ciconiidae Ciconia Ciconia ciconia                   NA   SPECIES
6 Ciconiidae Ciconia Ciconia ciconia                   NA   SPECIES
                   scientificName verbatimScientificName
1 Ciconia ciconia (Linnaeus, 1758)        Ciconia ciconia
2   Ciconia nigra (Linnaeus, 1758)          Ciconia nigra
3 Ciconia ciconia (Linnaeus, 1758)        Ciconia ciconia
4 Ciconia ciconia (Linnaeus, 1758)        Ciconia ciconia
5 Ciconia ciconia (Linnaeus, 1758)        Ciconia ciconia
6 Ciconia ciconia (Linnaeus, 1758)        Ciconia ciconia
  verbatimScientificNameAuthorship countryCode
```

```
1                               NA          PL
2                               NA          PL
3                               NA          PL
4                               NA          PL
5                               NA          PL
6                               NA          PL
                              locality  stateProvince occurrenceStatus
1                           Sandomierz Swietokrzyskie          PRESENT
2             Highway, southern Poland    Malopolskie          PRESENT
3 Białowieża National Park (IBA PL046)      Podlaskie          PRESENT
4              Biebrza NP--Dobarz area      Podlaskie          PRESENT
5            Reserwat Mierzeja Sarbska      Pomorskie          PRESENT
6                Poleski National Park      Lubelskie          PRESENT
  individualCount                     publishingOrgKey decimalLatitude
1              NA e2e717bf-551a-4917-bdc9-4fa0f342c530        50.67987
2               1 e2e717bf-551a-4917-bdc9-4fa0f342c530        49.64294
3              NA e2e717bf-551a-4917-bdc9-4fa0f342c530        52.77079
4              NA e2e717bf-551a-4917-bdc9-4fa0f342c530        53.36978
5               2 e2e717bf-551a-4917-bdc9-4fa0f342c530        54.77334
6               2 e2e717bf-551a-4917-bdc9-4fa0f342c530        51.37454
  decimalLongitude coordinateUncertaintyInMeters coordinatePrecision elevation
1         21.76563                            NA                  NA        NA
2         19.92021                            NA                  NA        NA
3         23.85561                            NA                  NA        NA
4         22.59019                            NA                  NA        NA
5         17.62782                            NA                  NA        NA
6         23.04717                            NA                  NA        NA
  elevationAccuracy depth depthAccuracy  eventDate day month year taxonKey
1                NA    NA            NA 2010-07-18  18     7 2010  2481912
2                NA    NA            NA 2010-07-21  21     7 2010  2481909
3                NA    NA            NA 2000-07-28  28     7 2000  2481912
4                NA    NA            NA 2000-07-28  28     7 2000  2481912
5                NA    NA            NA 2000-07-24  24     7 2000  2481912
6                NA    NA            NA 2010-07-30  30     7 2010  2481912
  speciesKey     basisOfRecord institutionCode collectionCode catalogNumber
1    2481912 HUMAN_OBSERVATION             CLO          EBIRD  OBS219042257
2    2481909 HUMAN_OBSERVATION             CLO          EBIRD  OBS207389323
3    2481912 HUMAN_OBSERVATION             CLO          EBIRD  OBS200802257
4    2481912 HUMAN_OBSERVATION             CLO          EBIRD  OBS200798945
5    2481912 HUMAN_OBSERVATION             CLO          EBIRD  OBS200758554
6    2481912 HUMAN_OBSERVATION             CLO          EBIRD   OBS96224279
  recordNumber identifiedBy dateIdentified   license rightsHolder recordedBy
1           NA           NA             NA CC_BY_4_0           NA obsr426265
```

```
2           NA           NA              NA CC_BY_4_0           NA obsr116419
3           NA           NA              NA CC_BY_4_0           NA obsr119031
4           NA           NA              NA CC_BY_4_0           NA obsr119031
5           NA           NA              NA CC_BY_4_0           NA obsr119031
6           NA           NA              NA CC_BY_4_0           NA obsr116419
  typeStatus establishmentMeans         lastInterpreted mediaType
1         NA                 NA 2024-04-17 08:23:22.915        NA
2         NA                 NA 2024-04-17 08:24:17.292        NA
3         NA                 NA 2024-04-17 09:07:13.585        NA
4         NA                 NA 2024-04-17 08:45:14.385        NA
5         NA                 NA 2024-04-17 09:51:22.303        NA
6         NA                 NA 2024-04-17 09:02:29.332        NA
                                                                     issue
1 CONTINENT_DERIVED_FROM_COORDINATES;TAXON_MATCH_TAXON_CONCEPT_ID_IGNORED
2 CONTINENT_DERIVED_FROM_COORDINATES;TAXON_MATCH_TAXON_CONCEPT_ID_IGNORED
3 CONTINENT_DERIVED_FROM_COORDINATES;TAXON_MATCH_TAXON_CONCEPT_ID_IGNORED
4 CONTINENT_DERIVED_FROM_COORDINATES;TAXON_MATCH_TAXON_CONCEPT_ID_IGNORED
5 CONTINENT_DERIVED_FROM_COORDINATES;TAXON_MATCH_TAXON_CONCEPT_ID_IGNORED
6 CONTINENT_DERIVED_FROM_COORDINATES;TAXON_MATCH_TAXON_CONCEPT_ID_IGNORED
  country
1  poland
2  poland
3  poland
4  poland
5  poland
6  poland
```

```
#596024662,584212444
```

info: all belong to aves class #important columns to keep: order, family, genus, species, stateProvince, individualCount, decimalLatitude, elevation,decimalLongitude, day,month, year, country

```
filtered_bird_orders <- bird_orders |>
  select(order, family, genus, species, stateProvince, individualCount, decimalLatitude, ele
head(filtered_bird_orders)
```

```
          order     family   genus         species  stateProvince
1 Ciconiiformes Ciconiidae Ciconia Ciconia ciconia Swietokrzyskie
2 Ciconiiformes Ciconiidae Ciconia   Ciconia nigra    Malopolskie
3 Ciconiiformes Ciconiidae Ciconia Ciconia ciconia      Podlaskie
4 Ciconiiformes Ciconiidae Ciconia Ciconia ciconia      Podlaskie
```

```
5 Ciconiiformes Ciconiidae Ciconia Ciconia ciconia      Pomorskie
6 Ciconiiformes Ciconiidae Ciconia Ciconia ciconia      Lubelskie
  individualCount decimalLatitude elevation decimalLongitude day month year
1              NA        50.67987        NA         21.76563  18     7 2010
2               1        49.64294        NA         19.92021  21     7 2010
3              NA        52.77079        NA         23.85561  28     7 2000
4              NA        53.36978        NA         22.59019  28     7 2000
5               2        54.77334        NA         17.62782  24     7 2000
6               2        51.37454        NA         23.04717  30     7 2010
  country
1  poland
2  poland
3  poland
4  poland
5  poland
6  poland
```

```r
country_mapping <- data.frame(
  iso = c("POL", "UKR", "URY", "IRL", "BGD", "TZA", "MOZ", "IDN", "COD", "PRY"),
  country = c("poland", "ukraine", "uruguay", "ireland", "bangladesh",
              "tanzania", "mozambique", "indonesia", "dcongo", "paraguay")
)

filtered_bird_orders$country <- tolower(filtered_bird_orders$country)

filtered_forest_data <- forest_data %>%
  filter(iso %in% country_mapping$iso) %>%   # Keep only rows with matching ISO codes
  left_join(country_mapping, by = "iso")

# Combine the datasets based on the "country" column
combined_data <- left_join(filtered_bird_orders, filtered_forest_data, by = "country")

# Output the combined dataset
write.csv(combined_data, "filtered_forest_data.csv")
```

```r
joined_data <- read.csv("filtered_forest_data.csv")
tail(joined_data)
```

```
          X         order      family    genus            species
92204 92204 Ciconiiformes Ciconiidae  Ciconia    Ciconia maguari
92205 92205 Ciconiiformes Ciconiidae  Ciconia    Ciconia maguari
92206 92206 Ciconiiformes Ciconiidae Mycteria Mycteria americana
```

```
92207 92207 Ciconiiformes Ciconiidae   Jabiru    Jabiru mycteria
92208 92208 Ciconiiformes Ciconiidae Mycteria Mycteria americana
92209 92209 Ciconiiformes Ciconiidae  Ciconia    Ciconia maguari
         stateProvince individualCount decimalLatitude elevation
92204        San Pedro              NA       -23.83000        NA
92205        San Pedro               2       -23.97583        NA
92206 Presidente Hayes               4       -22.55000        NA
92207 Presidente Hayes               1       -22.55000        NA
92208 Presidente Hayes              NA       -22.54028        NA
92209 Presidente Hayes              NA       -22.54028        NA
      decimalLongitude day month year  country iso   stable    loss     gain
92204        -56.16000   5     8 2000 paraguay PRY 14232500 5810873 642486.4
92205        -57.24650   6     8 2010 paraguay PRY 14232500 5810873 642486.4
92206        -59.38334  29     1 2010 paraguay PRY 14232500 5810873 642486.4
92207        -59.38334  29     1 2010 paraguay PRY 14232500 5810873 642486.4
92208        -59.67639   1    10 2000 paraguay PRY 14232500 5810873 642486.4
92209        -59.67639   1    10 2000 paraguay PRY 14232500 5810873 642486.4
       disturb      net   change gfw_area__ha
92204 883902.8 -5168387 -24.6969     39958592
92205 883902.8 -5168387 -24.6969     39958592
92206 883902.8 -5168387 -24.6969     39958592
92207 883902.8 -5168387 -24.6969     39958592
92208 883902.8 -5168387 -24.6969     39958592
92209 883902.8 -5168387 -24.6969     39958592
```

```r
# Aggregate bird count by order, country, and year
order_counts <- joined_data %>%
  group_by(order, country, year) %>%
  summarize(order_count = sum(individualCount, na.rm = TRUE), .groups = "drop")

# Merge aggregated counts back with deforestation data
model_data <- joined_data %>%
  select(country, year, iso, stable, loss, gain, disturb, net, change, gfw_area__ha) %>%
  distinct() %>%
  inner_join(order_counts, by = c("country", "year"))

head(model_data, 50)
```

```
      country year iso    stable      loss     gain   disturb      net   change
1      poland 2010 POL 9114498.5 485398.35 892077.3 975192.80 406678.9  3.84563
2      poland 2010 POL 9114498.5 485398.35 892077.3 975192.80 406678.9  3.84563
3      poland 2010 POL 9114498.5 485398.35 892077.3 975192.80 406678.9  3.84563
```

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4 | poland | 2010 | POL | 9114498.5 | 485398.35 | 892077.3 | 975192.80 | 406678.9 | 3.84563 |
| 5 | poland | 2000 | POL | 9114498.5 | 485398.35 | 892077.3 | 975192.80 | 406678.9 | 3.84563 |
| 6 | poland | 2000 | POL | 9114498.5 | 485398.35 | 892077.3 | 975192.80 | 406678.9 | 3.84563 |
| 7 | poland | 2000 | POL | 9114498.5 | 485398.35 | 892077.3 | 975192.80 | 406678.9 | 3.84563 |
| 8 | poland | 2000 | POL | 9114498.5 | 485398.35 | 892077.3 | 975192.80 | 406678.9 | 3.84563 |
| 9 | poland | 2020 | POL | 9114498.5 | 485398.35 | 892077.3 | 975192.80 | 406678.9 | 3.84563 |
| 10 | poland | 2020 | POL | 9114498.5 | 485398.35 | 892077.3 | 975192.80 | 406678.9 | 3.84563 |
| 11 | poland | 2020 | POL | 9114498.5 | 485398.35 | 892077.3 | 975192.80 | 406678.9 | 3.84563 |
| 12 | poland | 2020 | POL | 9114498.5 | 485398.35 | 892077.3 | 975192.80 | 406678.9 | 3.84563 |
| 13 | ukraine | 2010 | UKR | 9065233.1 | 499821.45 | 926367.2 | 895296.27 | 426545.8 | 4.07772 |
| 14 | ukraine | 2010 | UKR | 9065233.1 | 499821.45 | 926367.2 | 895296.27 | 426545.8 | 4.07772 |
| 15 | ukraine | 2010 | UKR | 9065233.1 | 499821.45 | 926367.2 | 895296.27 | 426545.8 | 4.07772 |
| 16 | ukraine | 2010 | UKR | 9065233.1 | 499821.45 | 926367.2 | 895296.27 | 426545.8 | 4.07772 |
| 17 | ukraine | 2020 | UKR | 9065233.1 | 499821.45 | 926367.2 | 895296.27 | 426545.8 | 4.07772 |
| 18 | ukraine | 2020 | UKR | 9065233.1 | 499821.45 | 926367.2 | 895296.27 | 426545.8 | 4.07772 |
| 19 | ukraine | 2020 | UKR | 9065233.1 | 499821.45 | 926367.2 | 895296.27 | 426545.8 | 4.07772 |
| 20 | ukraine | 2020 | UKR | 9065233.1 | 499821.45 | 926367.2 | 895296.27 | 426545.8 | 4.07772 |
| 21 | ukraine | 2000 | UKR | 9065233.1 | 499821.45 | 926367.2 | 895296.27 | 426545.8 | 4.07772 |
| 22 | uruguay | 2000 | URY | 659122.2 | 112760.42 | 654493.5 | 230107.58 | 541733.1 | 54.06570 |
| 23 | uruguay | 2000 | URY | 659122.2 | 112760.42 | 654493.5 | 230107.58 | 541733.1 | 54.06570 |
| 24 | uruguay | 2000 | URY | 659122.2 | 112760.42 | 654493.5 | 230107.58 | 541733.1 | 54.06570 |
| 25 | uruguay | 2000 | URY | 659122.2 | 112760.42 | 654493.5 | 230107.58 | 541733.1 | 54.06570 |
| 26 | uruguay | 2010 | URY | 659122.2 | 112760.42 | 654493.5 | 230107.58 | 541733.1 | 54.06570 |
| 27 | uruguay | 2010 | URY | 659122.2 | 112760.42 | 654493.5 | 230107.58 | 541733.1 | 54.06570 |
| 28 | uruguay | 2010 | URY | 659122.2 | 112760.42 | 654493.5 | 230107.58 | 541733.1 | 54.06570 |
| 29 | uruguay | 2010 | URY | 659122.2 | 112760.42 | 654493.5 | 230107.58 | 541733.1 | 54.06570 |
| 30 | uruguay | 2020 | URY | 659122.2 | 112760.42 | 654493.5 | 230107.58 | 541733.1 | 54.06570 |
| 31 | uruguay | 2020 | URY | 659122.2 | 112760.42 | 654493.5 | 230107.58 | 541733.1 | 54.06570 |
| 32 | uruguay | 2020 | URY | 659122.2 | 112760.42 | 654493.5 | 230107.58 | 541733.1 | 54.06570 |
| 33 | uruguay | 2020 | URY | 659122.2 | 112760.42 | 654493.5 | 230107.58 | 541733.1 | 54.06570 |
| 34 | ireland | 2010 | IRL | 388037.8 | 94639.68 | 284264.5 | 77309.85 | 189624.9 | 33.86240 |
| 35 | ireland | 2010 | IRL | 388037.8 | 94639.68 | 284264.5 | 77309.85 | 189624.9 | 33.86240 |
| 36 | ireland | 2010 | IRL | 388037.8 | 94639.68 | 284264.5 | 77309.85 | 189624.9 | 33.86240 |
| 37 | ireland | 2010 | IRL | 388037.8 | 94639.68 | 284264.5 | 77309.85 | 189624.9 | 33.86240 |
| 38 | ireland | 2000 | IRL | 388037.8 | 94639.68 | 284264.5 | 77309.85 | 189624.9 | 33.86240 |
| 39 | ireland | 2000 | IRL | 388037.8 | 94639.68 | 284264.5 | 77309.85 | 189624.9 | 33.86240 |
| 40 | ireland | 2000 | IRL | 388037.8 | 94639.68 | 284264.5 | 77309.85 | 189624.9 | 33.86240 |
| 41 | ireland | 2020 | IRL | 388037.8 | 94639.68 | 284264.5 | 77309.85 | 189624.9 | 33.86240 |
| 42 | ireland | 2020 | IRL | 388037.8 | 94639.68 | 284264.5 | 77309.85 | 189624.9 | 33.86240 |
| 43 | ireland | 2020 | IRL | 388037.8 | 94639.68 | 284264.5 | 77309.85 | 189624.9 | 33.86240 |
| 44 | bangladesh | 2020 | BGD | 3313032.5 | 115404.60 | 317862.2 | 612964.06 | 202457.6 | 5.00959 |
| 45 | bangladesh | 2020 | BGD | 3313032.5 | 115404.60 | 317862.2 | 612964.06 | 202457.6 | 5.00959 |
| 46 | bangladesh | 2020 | BGD | 3313032.5 | 115404.60 | 317862.2 | 612964.06 | 202457.6 | 5.00959 |

```
47 bangladesh 2020 BGD 3313032.5 115404.60 317862.2 612964.06 202457.6  5.00959
48 bangladesh 2010 BGD 3313032.5 115404.60 317862.2 612964.06 202457.6  5.00959
49 bangladesh 2010 BGD 3313032.5 115404.60 317862.2 612964.06 202457.6  5.00959
50 bangladesh 2010 BGD 3313032.5 115404.60 317862.2 612964.06 202457.6  5.00959
   gfw_area__ha         order order_count
1      31240006 Ciconiiformes         244
2      31240006 Coraciiformes          29
3      31240006   Galliformes         153
4      31240006  Strigiformes          22
5      31240006 Ciconiiformes         617
6      31240006 Coraciiformes          30
7      31240006   Galliformes          68
8      31240006  Strigiformes         191
9      31240006 Ciconiiformes        4857
10     31240006 Coraciiformes        2203
11     31240006   Galliformes        3273
12     31240006  Strigiformes         601
13     60136264 Ciconiiformes         165
14     60136264 Coraciiformes         129
15     60136264   Galliformes          15
16     60136264  Strigiformes           5
17     60136264 Ciconiiformes        5776
18     60136264 Coraciiformes        6836
19     60136264   Galliformes        2922
20     60136264  Strigiformes         826
21     60136264 Ciconiiformes         390
22     17746563 Ciconiiformes          10
23     17746563 Coraciiformes           2
24     17746563   Galliformes           0
25     17746563  Strigiformes          14
26     17746563 Ciconiiformes         122
27     17746563 Coraciiformes          25
28     17746563   Galliformes          19
29     17746563  Strigiformes          43
30     17746563 Ciconiiformes        3425
31     17746563 Coraciiformes        1623
32     17746563   Galliformes         686
33     17746563  Strigiformes        1437
34      7036677 Ciconiiformes           1
35      7036677 Coraciiformes          40
36      7036677   Galliformes          24
37      7036677  Strigiformes          15
38      7036677 Coraciiformes          13
```

```
39      7036677   Galliformes         14
40      7036677  Strigiformes          5
41      7036677 Coraciiformes        334
42      7036677   Galliformes       1132
43      7036677  Strigiformes         90
44     13938445 Ciconiiformes      13444
45     13938445 Coraciiformes      13264
46     13938445   Galliformes        271
47     13938445  Strigiformes        886
48     13938445 Ciconiiformes        324
49     13938445 Coraciiformes        121
50     13938445   Galliformes         12
```

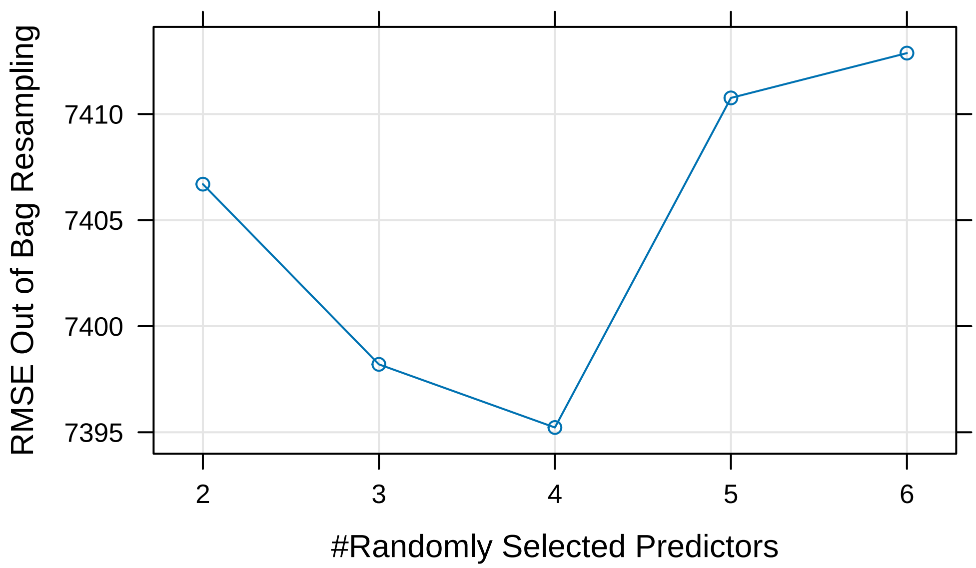
```r
set.seed(123)  # For reproducibility
library(caret)

# Split the data into training and testing sets
split_index <- createDataPartition(model_data$order_count, p = 0.8, list = FALSE)
train_data <- model_data[split_index, ]
test_data <- model_data[-split_index, ]
```

```r
library(randomForest)
library(caret)

# Random forest model with tuning
rf_model <- train(
  order_count ~ stable + loss + gain + disturb + net + change + gfw_area__ha,
  data = train_data,
  method = "rf",
  trControl = trainControl(method = "oob", search = "grid"),
  tuneGrid = expand.grid(mtry = 2:6)  # Test different values of mtry
)

# Print results
print(rf_model)
```

```
Random Forest

222 samples
  7 predictor

No pre-processing
```

```
Resampling results across tuning parameters:

  mtry  RMSE      Rsquared    
  2     7406.695  0.0019690806
  3     7398.202  0.0042567258
  4     7395.225  0.0050577667
  5     7410.761  0.0008730452
  6     7412.873  0.0003035085

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 4.
```

```
# Visualize the trained random forest model
plot(rf_model)
```




```
#calculate prediction accuracy for train and test data with the trained random forest
```

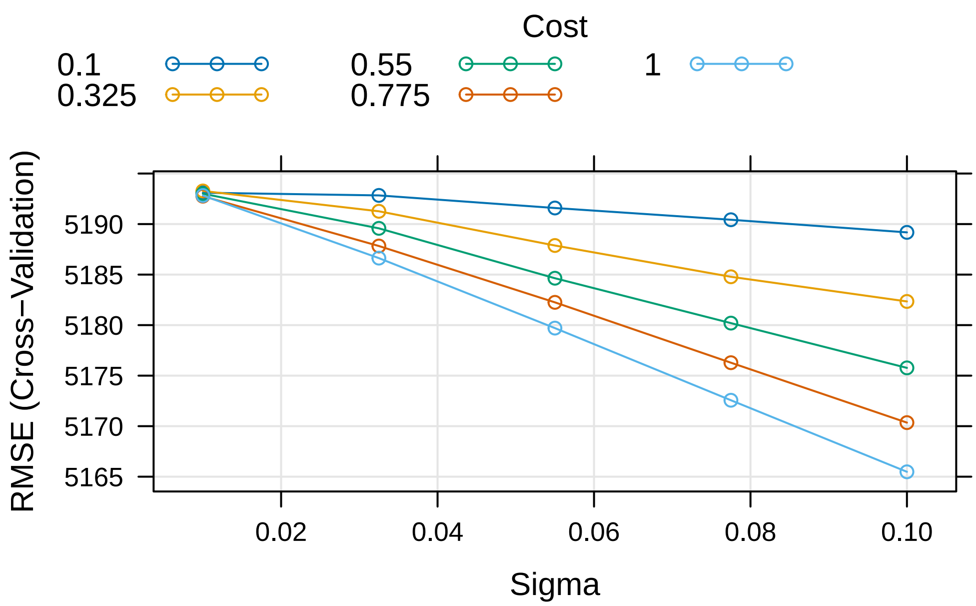
```
library(e1071)

# SVM model with tuning
svm_model <- train(
  order_count ~ stable + loss + gain + disturb + net + change + gfw_area__ha,
```

```r
  data = train_data,
  method = "svmRadial",
  trControl = trainControl(method = "cv", number = 10),  # 10-fold cross-validation
  tuneGrid = expand.grid(sigma = seq(0.01, 0.1, length.out = 5), C = seq(0.1, 1, length.out =
)

# Print results
print(svm_model)
```

```
Support Vector Machines with Radial Basis Function Kernel

222 samples
  7 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 200, 201, 199, 199, 201, 199, ...
Resampling results across tuning parameters:

  sigma   C      RMSE      Rsquared    MAE
  0.0100  0.100  5193.100  0.04018114  2107.596
  0.0100  0.325  5193.277  0.03981184  2108.696
  0.0100  0.550  5193.004  0.04461879  2109.582
  0.0100  0.775  5192.759  0.05404095  2110.365
  0.0100  1.000  5192.823  0.05325378  2111.190
  0.0325  0.100  5192.834  0.04116418  2107.789
  0.0325  0.325  5191.265  0.07461531  2109.733
  0.0325  0.550  5189.576  0.08402846  2111.327
  0.0325  0.775  5187.828  0.08816266  2111.616
  0.0325  1.000  5186.629  0.08987722  2112.388
  0.0550  0.100  5191.589  0.07087069  2107.861
  0.0550  0.325  5187.879  0.09220405  2108.424
  0.0550  0.550  5184.640  0.09511911  2108.904
  0.0550  0.775  5182.253  0.09502212  2109.094
  0.0550  1.000  5179.705  0.09637355  2109.390
  0.0775  0.100  5190.426  0.08277199  2107.188
  0.0775  0.325  5184.782  0.10469783  2107.142
  0.0775  0.550  5180.197  0.10263361  2106.774
  0.0775  0.775  5176.282  0.09960804  2106.473
  0.0775  1.000  5172.566  0.10097168  2106.297
  0.1000  0.100  5189.177  0.10033087  2107.029
  0.1000  0.325  5182.343  0.11313518  2106.548
```

```
  0.1000  0.550  5175.770  0.11342420  2104.673
  0.1000  0.775  5170.351  0.11031806  2104.505
  0.1000  1.000  5165.484  0.10745459  2104.500

RMSE was used to select the optimal model using the smallest value.
The final values used for the model were sigma = 0.1 and C = 1.
```

```r
# Visualize SVM model performance
plot(svm_model)
```

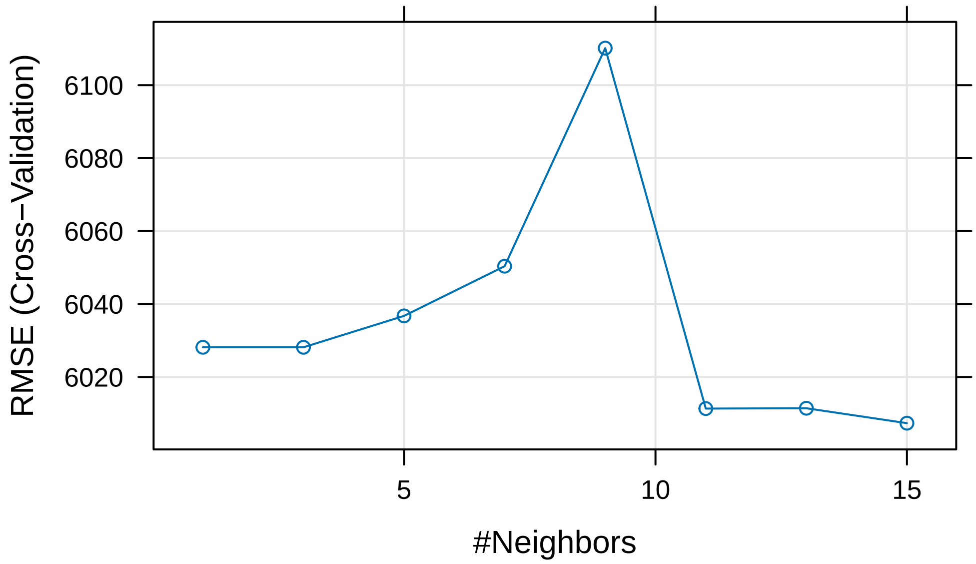



```r
#calculate prediction accuracy
```

```r
# KNN model with tuning
knn_model <- train(
  order_count ~ stable + loss + gain + disturb + net + change + gfw_area__ha,
  data = train_data,
  method = "knn",
  trControl = trainControl(method = "cv", number = 10),  # 10-fold cross-validation
  tuneGrid = expand.grid(k = seq(1, 15, by = 2))  # Test different values of k
)

# Print results
print(knn_model)
```

```
k-Nearest Neighbors

222 samples
  7 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 199, 200, 200, 201, 200, 201, ...
Resampling results across tuning parameters:

  k   RMSE      Rsquared    MAE
   1  6028.143  0.11998616  2723.384
   3  6028.143  0.11998616  2723.384
   5  6036.748  0.11881044  2741.087
   7  6050.376  0.11633222  2761.100
   9  6110.161  0.08989554  2781.701
  11  6011.341  0.05558459  2662.650
  13  6011.432  0.05555491  2663.248
  15  6007.338  0.05769225  2645.292

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was k = 15.
```

```r
# Visualize the KNN model performance
plot(knn_model)
```




```r
# Make predictions
rf_preds <- predict(rf_model, test_data)
knn_preds <- predict(knn_model, test_data)
svm_preds <- predict(svm_model, test_data)

# Evaluate performance (example: RMSE)
rf_rmse <- RMSE(rf_preds, test_data$order_count)
knn_rmse <- RMSE(knn_preds, test_data$order_count)
svm_rmse <- RMSE(svm_preds, test_data$order_count)

cat("Random Forest RMSE:", rf_rmse, "\n")
```

```
Random Forest RMSE: 4146.357
```

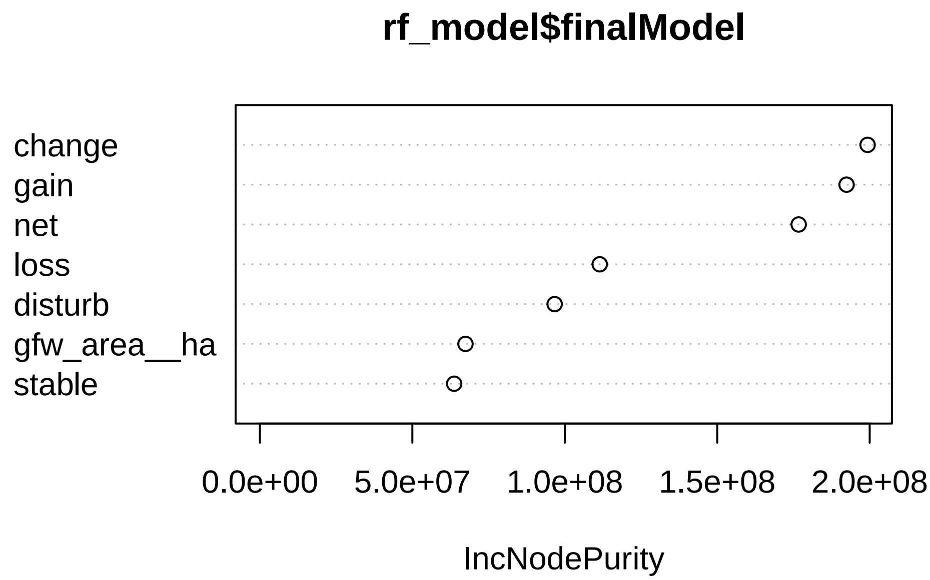
```r
cat("KNN RMSE:", knn_rmse, "\n")
```

```
KNN RMSE: 4057.214
```

```r
cat("SVM RMSE:", svm_rmse, "\n")
```

```
SVM RMSE: 4276.841
```

```r
# Variable importance plot
varImpPlot(rf_model$finalModel)
```

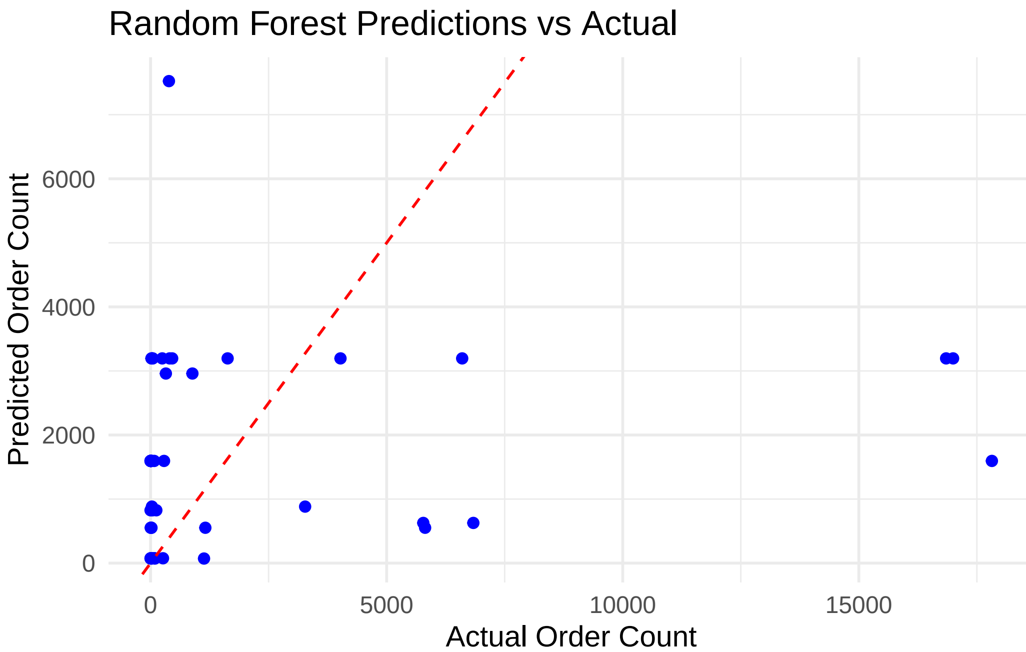



```r
# Predict order_count using each model
rf_preds <- predict(rf_model, test_data)  # Random Forest predictions
knn_preds <- predict(knn_model, test_data)  # KNN predictions
svm_preds <- predict(svm_model, test_data)  # SVM predictions

# Load necessary library for evaluation metrics
library(Metrics)

# Define metrics
calc_metrics <- function(actual, predicted) {
  r_squared <- cor(actual, predicted)^2
  rmse_val <- rmse(actual, predicted)
  mae_val <- mae(actual, predicted)

  return(data.frame(
    R_Squared = r_squared,
    RMSE = rmse_val,
    MAE = mae_val
  ))
```

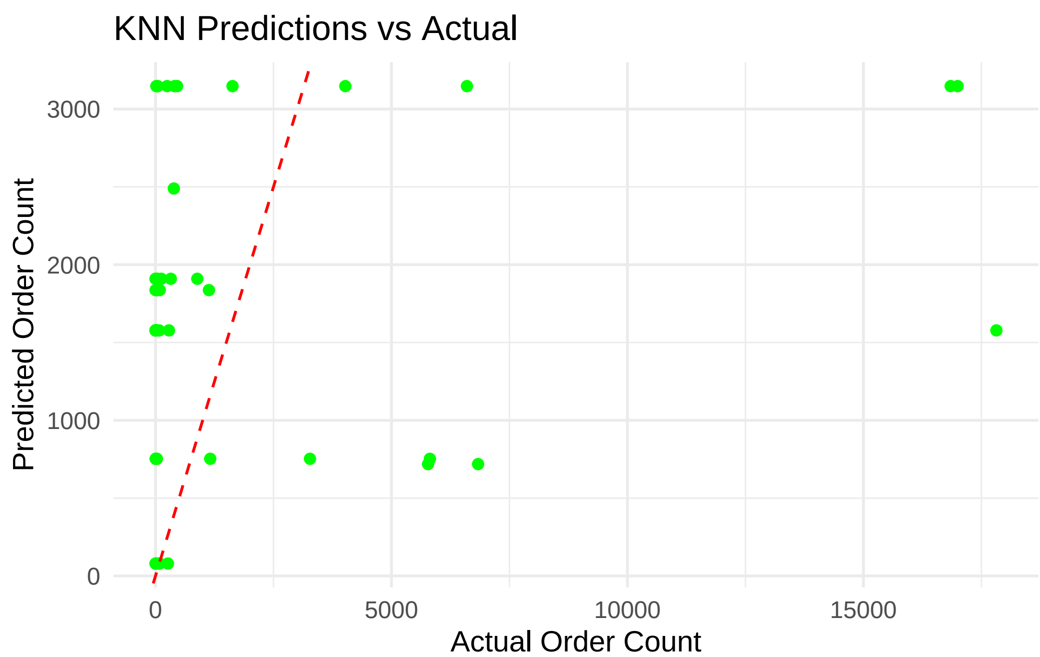
```r
}

# Calculate metrics for each model
rf_metrics <- calc_metrics(test_data$order_count, rf_preds)
knn_metrics <- calc_metrics(test_data$order_count, knn_preds)
svm_metrics <- calc_metrics(test_data$order_count, svm_preds)

# Combine results for comparison
model_metrics <- rbind(
  cbind(Model = "Random Forest", rf_metrics),
  cbind(Model = "KNN", knn_metrics),
  cbind(Model = "SVM", svm_metrics)
)

print(model_metrics)
```

```
          Model  R_Squared     RMSE      MAE
1 Random Forest 0.03483658 4146.357 2430.778
2           KNN 0.05433674 4057.214 2443.194
3           SVM 0.01165616 4276.841 1993.172
```

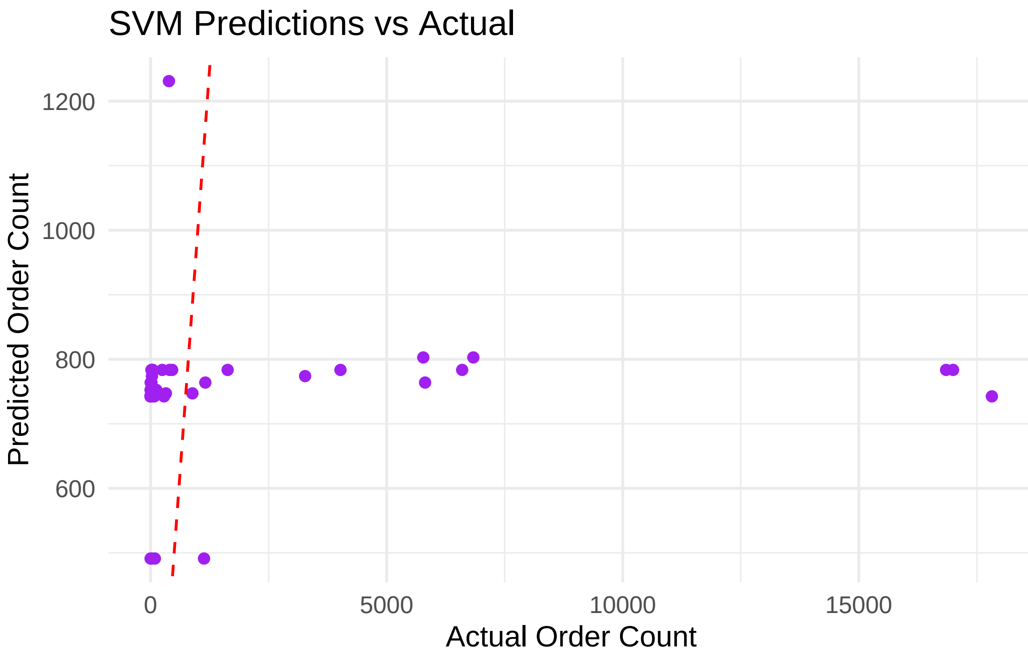
```r
# Combine actual and predicted data for visualization
test_results <- test_data %>%
  mutate(
    RF_Preds = rf_preds,
    KNN_Preds = knn_preds,
    SVM_Preds = svm_preds
  )

# Create scatter plots for each model
library(ggplot2)

ggplot(test_results, aes(x = order_count, y = RF_Preds)) +
  geom_point(color = "blue") +
  geom_abline(slope = 1, intercept = 0, linetype = "dashed", color = "red") +
  labs(
    title = "Random Forest Predictions vs Actual",
    x = "Actual Order Count",
    y = "Predicted Order Count"
  ) +
  theme_minimal()
```




```r
ggplot(test_results, aes(x = order_count, y = KNN_Preds)) +
  geom_point(color = "green") +
  geom_abline(slope = 1, intercept = 0, linetype = "dashed", color = "red") +
  labs(
    title = "KNN Predictions vs Actual",
    x = "Actual Order Count",
    y = "Predicted Order Count"
  ) +
  theme_minimal()
```




```r
ggplot(test_results, aes(x = order_count, y = SVM_Preds)) +
  geom_point(color = "purple") +
  geom_abline(slope = 1, intercept = 0, linetype = "dashed", color = "red") +
  labs(
    title = "SVM Predictions vs Actual",
    x = "Actual Order Count",
    y = "Predicted Order Count"
  ) +
  theme_minimal()
```



SVM Predictions vs Actual
Predicted Order Count
1200
1000
800
600
0
5000
10000
15000
Actual Order Count