PROJECT

2023-10-21

DATA PREPROCESSING

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ forcats 1.0.0 ✔ readr 2.1.4  
## ✔ ggplot2 3.4.2 ✔ stringr 1.5.0  
## ✔ lubridate 1.9.2 ✔ tibble 3.2.1  
## ✔ purrr 1.0.1 ✔ tidyr 1.3.0

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(ggplot2)

Importing Dataset

data<- read.csv('Bilateral\_Trade\_in\_Environmental\_Goods.csv')  
head(data)

## ObjectId Country ISO2 ISO3 Counterpart\_Country  
## 1 1 Afghanistan, Islamic Rep. of AF AFG Australia  
## 2 2 Afghanistan, Islamic Rep. of AF AFG Australia  
## 3 3 Afghanistan, Islamic Rep. of AF AFG Australia  
## 4 4 Afghanistan, Islamic Rep. of AF AFG Australia  
## 5 5 Afghanistan, Islamic Rep. of AF AFG Austria  
## 6 6 Afghanistan, Islamic Rep. of AF AFG Austria  
## Counterpart\_ISO2 Counterpart\_ISO3  
## 1 AU AUS  
## 2 AU AUS  
## 3 AU AUS  
## 4 AU AUS  
## 5 AT AUT  
## 6 AT AUT  
## Indicator Unit  
## 1 Environmental goods exports US Dollars  
## 2 Environmental goods exports as share of total exports Percent  
## 3 Environmental goods imports US Dollars  
## 4 Environmental goods imports as share of total imports Percent  
## 5 Environmental goods exports US Dollars  
## 6 Environmental goods exports as share of total exports Percent  
## Source  
## 1 Department of Economic and Social Affairs/United Nations. 2022. United Nations Comtrade database. https://comtrade.un.org. Accessed on 2023-06-28; IMF staff calculations.  
## 2 Department of Economic and Social Affairs/United Nations. 2022. United Nations Comtrade database. https://comtrade.un.org. Accessed on 2023-06-28; IMF staff calculations.  
## 3 Department of Economic and Social Affairs/United Nations. 2022. United Nations Comtrade database. https://comtrade.un.org. Accessed on 2023-06-28; IMF staff calculations.  
## 4 Department of Economic and Social Affairs/United Nations. 2022. United Nations Comtrade database. https://comtrade.un.org. Accessed on 2023-06-28; IMF staff calculations.  
## 5 Department of Economic and Social Affairs/United Nations. 2022. United Nations Comtrade database. https://comtrade.un.org. Accessed on 2023-06-28; IMF staff calculations.  
## 6 Department of Economic and Social Affairs/United Nations. 2022. United Nations Comtrade database. https://comtrade.un.org. Accessed on 2023-06-28; IMF staff calculations.  
## CTS\_Code CTS\_Name  
## 1 ECBTGX Trade in Environmental Goods; Exports  
## 2 ECBTGX Trade in Environmental Goods; Exports  
## 3 ECBTGM Trade in Environmental Goods; Imports  
## 4 ECBTGM Trade in Environmental Goods; Imports  
## 5 ECBTGX Trade in Environmental Goods; Exports  
## 6 ECBTGX Trade in Environmental Goods; Exports  
## CTS\_Full\_Descriptor  
## 1 Environment, Climate Change, Cross-Border, Trade-Related, Trade in Environmental Goods, Exports  
## 2 Environment, Climate Change, Cross-Border, Trade-Related, Trade in Environmental Goods, Exports  
## 3 Environment, Climate Change, Cross-Border, Trade-Related, Trade in Environmental Goods, Imports  
## 4 Environment, Climate Change, Cross-Border, Trade-Related, Trade in Environmental Goods, Imports  
## 5 Environment, Climate Change, Cross-Border, Trade-Related, Trade in Environmental Goods, Exports  
## 6 Environment, Climate Change, Cross-Border, Trade-Related, Trade in Environmental Goods, Exports  
## Trade\_Flow Scale F1994 F1995 F1996 F1997 F1998 F1999 F2000 F2001 F2002 F2003  
## 1 Exports Units NA NA NA NA NA NA NA NA NA NA  
## 2 Exports Units NA NA NA NA NA NA NA NA NA NA  
## 3 Imports Units NA NA NA NA NA NA NA NA NA NA  
## 4 Imports Units NA NA NA NA NA NA NA NA NA NA  
## 5 Exports Units NA NA NA NA NA NA NA NA NA NA  
## 6 Exports Units NA NA NA NA NA NA NA NA NA NA  
## F2004 F2005 F2006 F2007 F2008 F2009 F2010 F2011 F2012 F2013 F2014 F2015  
## 1 NA NA NA NA NA NA NA NA NA NA NA NA  
## 2 NA NA NA NA NA NA NA NA NA NA NA NA  
## 3 NA NA NA NA NA NA NA NA NA NA NA NA  
## 4 NA NA NA NA NA NA NA NA NA NA NA NA  
## 5 NA NA NA NA NA NA NA NA NA NA NA NA  
## 6 NA NA NA NA NA NA NA NA NA NA NA NA  
## F2016 F2017 F2018 F2019 F2020 F2021 F2022  
## 1 NA 145.95000000 NA NA NA NA NA  
## 2 NA 0.02430819 NA NA NA NA NA  
## 3 3.359500e+04 NA NA NA NA NA NA  
## 4 6.048676e-01 NA NA NA NA NA NA  
## 5 NA 729.43000000 658.4400000 NA NA NA NA  
## 6 NA 0.30453443 0.6306111 NA NA NA NA

Checking Shape of Dataset

shape <- dim(data)  
cat('Rows:', shape[1], 'Columns:', shape[2], '\n')

## Rows: 149801 Columns: 44

Checking data types of columns.

data\_types <- sapply(data, class)  
print(data\_types)

## ObjectId Country ISO2 ISO3   
## "integer" "character" "character" "character"   
## Counterpart\_Country Counterpart\_ISO2 Counterpart\_ISO3 Indicator   
## "character" "character" "character" "character"   
## Unit Source CTS\_Code CTS\_Name   
## "character" "character" "character" "character"   
## CTS\_Full\_Descriptor Trade\_Flow Scale F1994   
## "character" "character" "character" "numeric"   
## F1995 F1996 F1997 F1998   
## "numeric" "numeric" "numeric" "numeric"   
## F1999 F2000 F2001 F2002   
## "numeric" "numeric" "numeric" "numeric"   
## F2003 F2004 F2005 F2006   
## "numeric" "numeric" "numeric" "numeric"   
## F2007 F2008 F2009 F2010   
## "numeric" "numeric" "numeric" "numeric"   
## F2011 F2012 F2013 F2014   
## "numeric" "numeric" "numeric" "numeric"   
## F2015 F2016 F2017 F2018   
## "numeric" "numeric" "numeric" "numeric"   
## F2019 F2020 F2021 F2022   
## "numeric" "numeric" "numeric" "numeric"

Statistical Summary of Dataset

summary(data)

## ObjectId Country ISO2 ISO3   
## Min. : 1 Length:149801 Length:149801 Length:149801   
## 1st Qu.: 37451 Class :character Class :character Class :character   
## Median : 74901 Mode :character Mode :character Mode :character   
## Mean : 74957   
## 3rd Qu.:112351   
## Max. :153801   
##   
## Counterpart\_Country Counterpart\_ISO2 Counterpart\_ISO3 Indicator   
## Length:149801 Length:149801 Length:149801 Length:149801   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Unit Source CTS\_Code CTS\_Name   
## Length:149801 Length:149801 Length:149801 Length:149801   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## CTS\_Full\_Descriptor Trade\_Flow Scale F1994   
## Length:149801 Length:149801 Length:149801 Min. :-5.235e+09   
## Class :character Class :character Class :character 1st Qu.: 2.749e+04   
## Mode :character Mode :character Mode :character Median : 5.003e+05   
## Mean : 3.818e+07   
## 3rd Qu.: 5.805e+06   
## Max. : 1.342e+10   
## NA's :127054   
## F1995 F1996 F1997   
## Min. :-5.471e+09 Min. :-4.881e+09 Min. :-5.318e+09   
## 1st Qu.: 2.872e+04 1st Qu.: 3.027e+04 1st Qu.: 2.630e+04   
## Median : 5.292e+05 Median : 5.423e+05 Median : 4.937e+05   
## Mean : 4.276e+07 Mean : 4.198e+07 Mean : 4.145e+07   
## 3rd Qu.: 6.497e+06 3rd Qu.: 6.476e+06 3rd Qu.: 6.087e+06   
## Max. : 1.501e+10 Max. : 1.601e+10 Max. : 1.842e+10   
## NA's :124330 NA's :121992 NA's :120337   
## F1998 F1999 F2000   
## Min. :-5.417e+09 Min. :-5.793e+09 Min. :-6.848e+09   
## 1st Qu.: 4.000e+00 1st Qu.: 4.000e+00 1st Qu.: 3.000e+00   
## Median : 2.934e+04 Median : 2.152e+04 Median : 8.874e+03   
## Mean : 2.645e+07 Mean : 2.625e+07 Mean : 2.484e+07   
## 3rd Qu.: 1.778e+06 3rd Qu.: 1.447e+06 3rd Qu.: 9.393e+05   
## Max. : 1.977e+10 Max. : 2.187e+10 Max. : 2.475e+10   
## NA's :102472 NA's :99841 NA's :90057   
## F2001 F2002 F2003   
## Min. :-6.441e+09 Min. :-5.942e+09 Min. :-9.733e+09   
## 1st Qu.: 3.000e+00 1st Qu.: 3.000e+00 1st Qu.: 3.000e+00   
## Median : 8.283e+03 Median : 8.225e+03 Median : 8.818e+03   
## Mean : 2.375e+07 Mean : 2.401e+07 Mean : 2.759e+07   
## 3rd Qu.: 9.689e+05 3rd Qu.: 9.342e+05 3rd Qu.: 1.035e+06   
## Max. : 2.353e+10 Max. : 2.298e+10 Max. : 2.331e+10   
## NA's :88094 NA's :86005 NA's :84074   
## F2004 F2005 F2006   
## Min. :-1.367e+10 Min. :-1.328e+10 Min. :-1.676e+10   
## 1st Qu.: 3.000e+00 1st Qu.: 3.000e+00 1st Qu.: 3.000e+00   
## Median : 9.208e+03 Median : 1.000e+04 Median : 9.014e+03   
## Mean : 3.314e+07 Mean : 3.628e+07 Mean : 4.058e+07   
## 3rd Qu.: 1.160e+06 3rd Qu.: 1.290e+06 3rd Qu.: 1.382e+06   
## Max. : 2.547e+10 Max. : 2.771e+10 Max. : 3.019e+10   
## NA's :81671 NA's :79590 NA's :77087   
## F2007 F2008 F2009   
## Min. :-2.000e+10 Min. :-2.088e+10 Min. :-1.608e+10   
## 1st Qu.: 3.000e+00 1st Qu.: 4.000e+00 1st Qu.: 4.000e+00   
## Median : 1.035e+04 Median : 1.104e+04 Median : 1.019e+04   
## Mean : 4.655e+07 Mean : 5.276e+07 Mean : 4.326e+07   
## 3rd Qu.: 1.657e+06 3rd Qu.: 1.982e+06 3rd Qu.: 1.645e+06   
## Max. : 3.221e+10 Max. : 3.415e+10 Max. : 2.790e+10   
## NA's :74143 NA's :72652 NA's :72747   
## F2010 F2011 F2012   
## Min. :-2.277e+10 Min. :-2.414e+10 Min. :-2.259e+10   
## 1st Qu.: 4.000e+00 1st Qu.: 4.000e+00 1st Qu.: 4.000e+00   
## Median : 9.955e+03 Median : 1.113e+04 Median : 1.108e+04   
## Mean : 5.052e+07 Mean : 5.797e+07 Mean : 5.829e+07   
## 3rd Qu.: 1.716e+06 3rd Qu.: 1.975e+06 3rd Qu.: 1.963e+06   
## Max. : 3.780e+10 Max. : 4.345e+10 Max. : 4.499e+10   
## NA's :69883 NA's :69813 NA's :67996   
## F2013 F2014 F2015   
## Min. :-2.213e+10 Min. :-2.125e+10 Min. :-2.416e+10   
## 1st Qu.: 4.000e+00 1st Qu.: 4.000e+00 1st Qu.: 4.000e+00   
## Median : 1.124e+04 Median : 1.096e+04 Median : 9.185e+03   
## Mean : 5.803e+07 Mean : 5.989e+07 Mean : 5.392e+07   
## 3rd Qu.: 2.020e+06 3rd Qu.: 2.115e+06 3rd Qu.: 1.781e+06   
## Max. : 4.521e+10 Max. : 4.966e+10 Max. : 5.013e+10   
## NA's :66098 NA's :66472 NA's :63899   
## F2016 F2017 F2018   
## Min. :-2.368e+10 Min. :-2.468e+10 Min. :-2.931e+10   
## 1st Qu.: 4.000e+00 1st Qu.: 4.000e+00 1st Qu.: 4.000e+00   
## Median : 8.131e+03 Median : 8.338e+03 Median : 9.636e+03   
## Mean : 5.221e+07 Mean : 5.773e+07 Mean : 6.302e+07   
## 3rd Qu.: 1.629e+06 3rd Qu.: 1.737e+06 3rd Qu.: 1.921e+06   
## Max. : 4.974e+10 Max. : 5.402e+10 Max. : 5.727e+10   
## NA's :62569 NA's :61109 NA's :61652   
## F2019 F2020 F2021   
## Min. :-2.399e+10 Min. :-2.533e+10 Min. :-3.305e+10   
## 1st Qu.: 4.000e+00 1st Qu.: 5.000e+00 1st Qu.: 5.000e+00   
## Median : 9.410e+03 Median : 9.944e+03 Median : 1.135e+04   
## Mean : 6.368e+07 Mean : 6.580e+07 Mean : 8.339e+07   
## 3rd Qu.: 1.874e+06 3rd Qu.: 1.919e+06 3rd Qu.: 2.206e+06   
## Max. : 5.691e+10 Max. : 5.542e+10 Max. : 6.957e+10   
## NA's :61751 NA's :64425 NA's :65980   
## F2022   
## Min. :-1.241e+10   
## 1st Qu.: 4.000e+00   
## Median : 1.052e+04   
## Mean : 6.217e+07   
## 3rd Qu.: 1.958e+06   
## Max. : 6.594e+10   
## NA's :118829

Checking null values

null\_values\_count<- colSums(is.na(data))  
null\_values\_dataframe <- data.frame(Column = names(null\_values\_count), NA\_Count = null\_values\_count)  
null\_values\_dataframe

## Column NA\_Count  
## ObjectId ObjectId 0  
## Country Country 0  
## ISO2 ISO2 847  
## ISO3 ISO3 0  
## Counterpart\_Country Counterpart\_Country 0  
## Counterpart\_ISO2 Counterpart\_ISO2 704  
## Counterpart\_ISO3 Counterpart\_ISO3 0  
## Indicator Indicator 0  
## Unit Unit 0  
## Source Source 0  
## CTS\_Code CTS\_Code 0  
## CTS\_Name CTS\_Name 0  
## CTS\_Full\_Descriptor CTS\_Full\_Descriptor 0  
## Trade\_Flow Trade\_Flow 0  
## Scale Scale 0  
## F1994 F1994 127054  
## F1995 F1995 124330  
## F1996 F1996 121992  
## F1997 F1997 120337  
## F1998 F1998 102472  
## F1999 F1999 99841  
## F2000 F2000 90057  
## F2001 F2001 88094  
## F2002 F2002 86005  
## F2003 F2003 84074  
## F2004 F2004 81671  
## F2005 F2005 79590  
## F2006 F2006 77087  
## F2007 F2007 74143  
## F2008 F2008 72652  
## F2009 F2009 72747  
## F2010 F2010 69883  
## F2011 F2011 69813  
## F2012 F2012 67996  
## F2013 F2013 66098  
## F2014 F2014 66472  
## F2015 F2015 63899  
## F2016 F2016 62569  
## F2017 F2017 61109  
## F2018 F2018 61652  
## F2019 F2019 61751  
## F2020 F2020 64425  
## F2021 F2021 65980  
## F2022 F2022 118829

Removing Some columns from dataset.

columns\_to\_remove <- c('ISO3', 'Counterpart\_ISO3', 'Source', 'CTS\_Code', 'CTS\_Name', 'CTS\_Full\_Descriptor', 'Unit', 'Scale', 'F1994', 'F1995', 'F1996', 'F1997', 'F1998', 'F1999', 'F2000', 'F2001', 'F2002', 'F2003', 'F2004', 'F2005', 'F2006', 'F2007', 'F2008', 'F2009', 'F2010', 'F2011')  
  
filtered\_data <- data[, !(names(data) %in% columns\_to\_remove)]  
head(filtered\_data,2)

## ObjectId Country ISO2 Counterpart\_Country  
## 1 1 Afghanistan, Islamic Rep. of AF Australia  
## 2 2 Afghanistan, Islamic Rep. of AF Australia  
## Counterpart\_ISO2 Indicator  
## 1 AU Environmental goods exports  
## 2 AU Environmental goods exports as share of total exports  
## Trade\_Flow F2012 F2013 F2014 F2015 F2016 F2017 F2018 F2019 F2020 F2021  
## 1 Exports NA NA NA NA NA 145.95000000 NA NA NA NA  
## 2 Exports NA NA NA NA NA 0.02430819 NA NA NA NA  
## F2022  
## 1 NA  
## 2 NA

extracting columns of filtered Data.

names(filtered\_data)

## [1] "ObjectId" "Country" "ISO2"   
## [4] "Counterpart\_Country" "Counterpart\_ISO2" "Indicator"   
## [7] "Trade\_Flow" "F2012" "F2013"   
## [10] "F2014" "F2015" "F2016"   
## [13] "F2017" "F2018" "F2019"   
## [16] "F2020" "F2021" "F2022"

unique\_country <- table(data$Country)  
cat('There are',length(unique\_country),'unique countries in dataset')

## There are 208 unique countries in dataset

Checking null values of filtered\_data

null\_values\_count<- colSums(is.na(filtered\_data))  
null\_values\_dataframe <- data.frame(Column = names(null\_values\_count), NA\_Count = null\_values\_count)  
null\_values\_dataframe

## Column NA\_Count  
## ObjectId ObjectId 0  
## Country Country 0  
## ISO2 ISO2 847  
## Counterpart\_Country Counterpart\_Country 0  
## Counterpart\_ISO2 Counterpart\_ISO2 704  
## Indicator Indicator 0  
## Trade\_Flow Trade\_Flow 0  
## F2012 F2012 67996  
## F2013 F2013 66098  
## F2014 F2014 66472  
## F2015 F2015 63899  
## F2016 F2016 62569  
## F2017 F2017 61109  
## F2018 F2018 61652  
## F2019 F2019 61751  
## F2020 F2020 64425  
## F2021 F2021 65980  
## F2022 F2022 118829

Droping null values

data\_cleaned <- na.omit(filtered\_data)

Checking shape of cleaned data

shape <- dim(data\_cleaned)  
cat('Rows:', shape[1], 'Columns:', shape[2], '\n')

## Rows: 19741 Columns: 18

head(data\_cleaned)

## ObjectId Country ISO2 Counterpart\_Country  
## 1337 1337 Andorra, Principality of AD Andorra, Principality of  
## 1352 1352 Andorra, Principality of AD Austria  
## 1364 1364 Andorra, Principality of AD Belgium  
## 1365 1365 Andorra, Principality of AD Belgium  
## 1366 1366 Andorra, Principality of AD Belgium  
## 1367 1367 Andorra, Principality of AD Belgium  
## Counterpart\_ISO2 Indicator Trade\_Flow  
## 1337 AD Environmental goods imports Imports  
## 1352 AT Environmental goods imports Imports  
## 1364 BE Environmental goods exports Exports  
## 1365 BE Environmental goods imports Imports  
## 1366 BE Environmental goods trade balance Not Applicable  
## 1367 BE Total trade in environmental goods Not Applicable  
## F2012 F2013 F2014 F2015 F2016 F2017  
## 1337 124176.704 72869.57 15552.94 688.099 408.981 8042.192  
## 1352 57436.525 79987.58 616090.09 69688.266 123826.566 266202.777  
## 1364 4977.802 17650.64 23345.22 11296.396 19243.477 5714.549  
## 1365 55250.193 64066.09 140259.66 52245.280 124409.763 66914.236  
## 1366 -50272.391 -46415.45 -116914.44 -40948.884 -105166.286 -61199.687  
## 1367 60227.995 81716.73 163604.87 63541.676 143653.240 72628.785  
## F2018 F2019 F2020 F2021 F2022  
## 1337 4921.683 831.419 9112.604 6622.224 12337.32  
## 1352 646338.649 471439.576 456033.149 1264195.860 311349.64  
## 1364 12698.171 73595.050 1090.393 3052.548 347615.58  
## 1365 80759.229 64025.268 45854.805 168176.728 173184.81  
## 1366 -68061.058 9569.782 -44764.412 -165124.180 174430.77  
## 1367 93457.400 137620.318 46945.198 171229.276 520800.40

Changing value of object id and resetting index

data\_cleaned$ObjectId <- seq(1, nrow(data\_cleaned))  
head(data\_cleaned)

## ObjectId Country ISO2 Counterpart\_Country  
## 1337 1 Andorra, Principality of AD Andorra, Principality of  
## 1352 2 Andorra, Principality of AD Austria  
## 1364 3 Andorra, Principality of AD Belgium  
## 1365 4 Andorra, Principality of AD Belgium  
## 1366 5 Andorra, Principality of AD Belgium  
## 1367 6 Andorra, Principality of AD Belgium  
## Counterpart\_ISO2 Indicator Trade\_Flow  
## 1337 AD Environmental goods imports Imports  
## 1352 AT Environmental goods imports Imports  
## 1364 BE Environmental goods exports Exports  
## 1365 BE Environmental goods imports Imports  
## 1366 BE Environmental goods trade balance Not Applicable  
## 1367 BE Total trade in environmental goods Not Applicable  
## F2012 F2013 F2014 F2015 F2016 F2017  
## 1337 124176.704 72869.57 15552.94 688.099 408.981 8042.192  
## 1352 57436.525 79987.58 616090.09 69688.266 123826.566 266202.777  
## 1364 4977.802 17650.64 23345.22 11296.396 19243.477 5714.549  
## 1365 55250.193 64066.09 140259.66 52245.280 124409.763 66914.236  
## 1366 -50272.391 -46415.45 -116914.44 -40948.884 -105166.286 -61199.687  
## 1367 60227.995 81716.73 163604.87 63541.676 143653.240 72628.785  
## F2018 F2019 F2020 F2021 F2022  
## 1337 4921.683 831.419 9112.604 6622.224 12337.32  
## 1352 646338.649 471439.576 456033.149 1264195.860 311349.64  
## 1364 12698.171 73595.050 1090.393 3052.548 347615.58  
## 1365 80759.229 64025.268 45854.805 168176.728 173184.81  
## 1366 -68061.058 9569.782 -44764.412 -165124.180 174430.77  
## 1367 93457.400 137620.318 46945.198 171229.276 520800.40

save filtered data into csv file

write.csv(data\_cleaned, "Data.csv", row.names = FALSE)

Filter dataset

data<-read.csv("Data.csv")  
head(data)

## ObjectId Country ISO2 Counterpart\_Country  
## 1 1 Andorra, Principality of AD Andorra, Principality of  
## 2 2 Andorra, Principality of AD Austria  
## 3 3 Andorra, Principality of AD Belgium  
## 4 4 Andorra, Principality of AD Belgium  
## 5 5 Andorra, Principality of AD Belgium  
## 6 6 Andorra, Principality of AD Belgium  
## Counterpart\_ISO2 Indicator Trade\_Flow F2012  
## 1 AD Environmental goods imports Imports 124176.704  
## 2 AT Environmental goods imports Imports 57436.525  
## 3 BE Environmental goods exports Exports 4977.802  
## 4 BE Environmental goods imports Imports 55250.193  
## 5 BE Environmental goods trade balance Not Applicable -50272.391  
## 6 BE Total trade in environmental goods Not Applicable 60227.995  
## F2013 F2014 F2015 F2016 F2017 F2018 F2019  
## 1 72869.57 15552.94 688.099 408.981 8042.192 4921.683 831.419  
## 2 79987.58 616090.09 69688.266 123826.566 266202.777 646338.649 471439.576  
## 3 17650.64 23345.22 11296.396 19243.477 5714.549 12698.171 73595.050  
## 4 64066.09 140259.66 52245.280 124409.763 66914.236 80759.229 64025.268  
## 5 -46415.45 -116914.44 -40948.884 -105166.286 -61199.687 -68061.058 9569.782  
## 6 81716.73 163604.87 63541.676 143653.240 72628.785 93457.400 137620.318  
## F2020 F2021 F2022  
## 1 9112.604 6622.224 12337.32  
## 2 456033.149 1264195.860 311349.64  
## 3 1090.393 3052.548 347615.58  
## 4 45854.805 168176.728 173184.81  
## 5 -44764.412 -165124.180 174430.77  
## 6 46945.198 171229.276 520800.40

No. of unique countries after filtering dataset

unique\_country <- table(data$Country)  
cat('There are',length(unique\_country),'unique countries in dataset')

## There are 48 unique countries in dataset

Exploratory Data Analysis

— SMART QUESTION 01: What is the trend in global environmental goods exports over the last decade (2013-2022)?

exports\_data <- filter(data, Indicator == "Environmental goods exports")   
head(exports\_data)

## ObjectId Country ISO2 Counterpart\_Country Counterpart\_ISO2  
## 1 3 Andorra, Principality of AD Belgium BE  
## 2 14 Andorra, Principality of AD France FR  
## 3 18 Andorra, Principality of AD Germany DE  
## 4 26 Andorra, Principality of AD Italy IT  
## 5 35 Andorra, Principality of AD Netherlands, The NL  
## 6 39 Andorra, Principality of AD Norway NO  
## Indicator Trade\_Flow F2012 F2013 F2014  
## 1 Environmental goods exports Exports 4977.802 17650.636 23345.22  
## 2 Environmental goods exports Exports 333741.685 654987.125 856655.32  
## 3 Environmental goods exports Exports 17916.401 16433.645 42109.43  
## 4 Environmental goods exports Exports 58748.693 195708.618 231083.88  
## 5 Environmental goods exports Exports 18278.694 9245.024 31706.86  
## 6 Environmental goods exports Exports 45672.751 40555.439 46440.88  
## F2015 F2016 F2017 F2018 F2019 F2020  
## 1 11296.40 19243.477 5714.549 12698.171 73595.050 1090.393  
## 2 1030334.07 1234084.840 1016697.537 1100321.851 1356900.374 2421820.712  
## 3 16199.17 36896.948 22514.621 1001.678 2921.677 12279.488  
## 4 86439.44 181035.927 483990.169 8248.304 2245.843 2856.204  
## 5 10940.51 8314.891 7467.299 2022.019 119551.294 27688.404  
## 6 15371.52 4304.702 5937.560 2805.334 2446.793 9370.652  
## F2021 F2022  
## 1 3052.548 347615.58  
## 2 2450100.158 2353566.87  
## 3 633214.245 142628.66  
## 4 7094.110 35013.62  
## 5 1176.319 23766.57  
## 6 86769.964 65582.46

exports\_trend <- colSums(exports\_data[, c("F2013", "F2014", "F2015", "F2016", "F2017", "F2018", "F2019", "F2020", "F2021", "F2022")], na.rm = TRUE)  
exports\_trend

## F2013 F2014 F2015 F2016 F2017 F2018   
## 334665669498 344215245140 308304010583 313092795139 362823189353 394046482423   
## F2019 F2020 F2021 F2022   
## 392301626678 389186528113 475843460975 484885451576

exports\_trend <- round(exports\_trend / 1000000000) #Converting the tota cost of exports into billion

exports\_trend\_df <- data.frame(Year = 2013:2022, 'Exports\_In\_Billions' = exports\_trend)  
head(exports\_trend\_df)

## Year Exports\_In\_Billions  
## F2013 2013 335  
## F2014 2014 344  
## F2015 2015 308  
## F2016 2016 313  
## F2017 2017 363  
## F2018 2018 394

ggplot(exports\_trend\_df, aes(x = Year, y = Exports\_In\_Billions)) +  
 geom\_line(group = 1, color = '#38599F') +  
 geom\_point(color="#00297b") +  
 labs(title = "Global Environmental Goods Exports (2013-2022)",  
 x = "Year",  
 y = "Total Exports (US Dollars in billions)") +  
   
 scale\_x\_continuous(breaks = 2013:2022) +  
 scale\_y\_continuous(breaks = seq(300, 500, by = 25))

 >Linear Regression Model

env\_goods\_exports <- subset(data, Indicator == 'Environmental goods exports')  
head(env\_goods\_exports)

## ObjectId Country ISO2 Counterpart\_Country Counterpart\_ISO2  
## 3 3 Andorra, Principality of AD Belgium BE  
## 14 14 Andorra, Principality of AD France FR  
## 18 18 Andorra, Principality of AD Germany DE  
## 26 26 Andorra, Principality of AD Italy IT  
## 35 35 Andorra, Principality of AD Netherlands, The NL  
## 39 39 Andorra, Principality of AD Norway NO  
## Indicator Trade\_Flow F2012 F2013 F2014  
## 3 Environmental goods exports Exports 4977.802 17650.636 23345.22  
## 14 Environmental goods exports Exports 333741.685 654987.125 856655.32  
## 18 Environmental goods exports Exports 17916.401 16433.645 42109.43  
## 26 Environmental goods exports Exports 58748.693 195708.618 231083.88  
## 35 Environmental goods exports Exports 18278.694 9245.024 31706.86  
## 39 Environmental goods exports Exports 45672.751 40555.439 46440.88  
## F2015 F2016 F2017 F2018 F2019 F2020  
## 3 11296.40 19243.477 5714.549 12698.171 73595.050 1090.393  
## 14 1030334.07 1234084.840 1016697.537 1100321.851 1356900.374 2421820.712  
## 18 16199.17 36896.948 22514.621 1001.678 2921.677 12279.488  
## 26 86439.44 181035.927 483990.169 8248.304 2245.843 2856.204  
## 35 10940.51 8314.891 7467.299 2022.019 119551.294 27688.404  
## 39 15371.52 4304.702 5937.560 2805.334 2446.793 9370.652  
## F2021 F2022  
## 3 3052.548 347615.58  
## 14 2450100.158 2353566.87  
## 18 633214.245 142628.66  
## 26 7094.110 35013.62  
## 35 1176.319 23766.57  
## 39 86769.964 65582.46

years <- paste0('F', 2013:2022)  
env\_goods\_exports\_years <- env\_goods\_exports[, c('Country', years)]  
head(env\_goods\_exports\_years)

## Country F2013 F2014 F2015 F2016  
## 3 Andorra, Principality of 17650.636 23345.22 11296.40 19243.477  
## 14 Andorra, Principality of 654987.125 856655.32 1030334.07 1234084.840  
## 18 Andorra, Principality of 16433.645 42109.43 16199.17 36896.948  
## 26 Andorra, Principality of 195708.618 231083.88 86439.44 181035.927  
## 35 Andorra, Principality of 9245.024 31706.86 10940.51 8314.891  
## 39 Andorra, Principality of 40555.439 46440.88 15371.52 4304.702  
## F2017 F2018 F2019 F2020 F2021 F2022  
## 3 5714.549 12698.171 73595.050 1090.393 3052.548 347615.58  
## 14 1016697.537 1100321.851 1356900.374 2421820.712 2450100.158 2353566.87  
## 18 22514.621 1001.678 2921.677 12279.488 633214.245 142628.66  
## 26 483990.169 8248.304 2245.843 2856.204 7094.110 35013.62  
## 35 7467.299 2022.019 119551.294 27688.404 1176.319 23766.57  
## 39 5937.560 2805.334 2446.793 9370.652 86769.964 65582.46

total\_exports\_by\_year <- colSums(env\_goods\_exports\_years[, -1], na.rm = TRUE)  
head(total\_exports\_by\_year, 3)

## F2013 F2014 F2015   
## 334665669498 344215245140 308304010583

total\_exports\_df <- data.frame(Year = as.numeric(sub("F", "", names(total\_exports\_by\_year))),   
 Total\_Export\_Value = total\_exports\_by\_year)  
total\_exports\_df <- arrange(total\_exports\_df, Year)  
total\_exports\_df$Total\_Export\_Value <- (total\_exports\_df$Total\_Export\_Value/1000000000)  
head(total\_exports\_df)

## Year Total\_Export\_Value  
## F2013 2013 334.6657  
## F2014 2014 344.2152  
## F2015 2015 308.3040  
## F2016 2016 313.0928  
## F2017 2017 362.8232  
## F2018 2018 394.0465

variance\_exports <- var(total\_exports\_df$Total\_Export\_Value)  
sd\_exports <- sd(total\_exports\_df$Total\_Export\_Value)

variance\_exports <- var(total\_exports\_df$Total\_Export\_Value)  
sd\_exports <- sd(total\_exports\_df$Total\_Export\_Value)  
  
cat('Variance: ',variance\_exports, 'Standard deviation: ',sd\_exports)

## Variance: 3763.066 Standard deviation: 61.34383

print(total\_exports\_df)

## Year Total\_Export\_Value  
## F2013 2013 334.6657  
## F2014 2014 344.2152  
## F2015 2015 308.3040  
## F2016 2016 313.0928  
## F2017 2017 362.8232  
## F2018 2018 394.0465  
## F2019 2019 392.3016  
## F2020 2020 389.1865  
## F2021 2021 475.8435  
## F2022 2022 484.8855

cat('Variance of Exports:', variance\_exports, '\n')

## Variance of Exports: 3763.066

cat('Standard Deviation of Exports:', sd\_exports, '\n')

## Standard Deviation of Exports: 61.34383

linear\_model <- lm(Total\_Export\_Value ~ Year, data = total\_exports\_df)  
model\_summary <- summary(linear\_model)

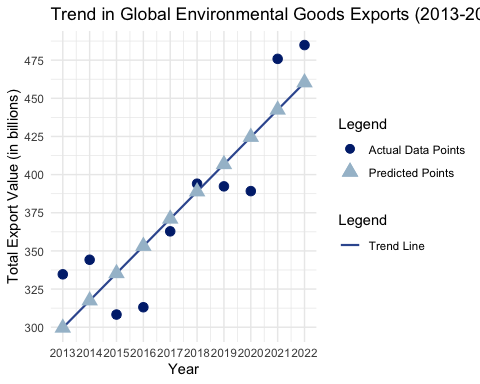
head(total\_exports\_df)

## Year Total\_Export\_Value  
## F2013 2013 334.6657  
## F2014 2014 344.2152  
## F2015 2015 308.3040  
## F2016 2016 313.0928  
## F2017 2017 362.8232  
## F2018 2018 394.0465

total\_exports\_df$residuals <- residuals(linear\_model)  
total\_exports\_df$predicted <- fitted(linear\_model)  
  
plot <- ggplot(total\_exports\_df, aes(x = Year)) +  
 geom\_point(aes(y = Total\_Export\_Value, color = "Actual Data Points"), size = 3) +  
 geom\_line(aes(y = predicted, linetype = "Trend Line"), size = 0.77, color = "#38599f") +  
 geom\_point(aes(y = predicted, color = "Predicted Points"), size = 4, shape = 17) +  
 scale\_color\_manual(values = c("Actual Data Points" = "#00297b",   
 "Predicted Points" = "#a5bfd1")) +  
 scale\_linetype\_manual(values = c("Trend Line" = "solid")) +  
 labs(title = "Trend in Global Environmental Goods Exports (2013-2022)",  
 x = "Year",  
 y = "Total Export Value (in billions)",  
 color = "Legend",  
 linetype = "Legend") +  
 theme\_minimal() +  
 scale\_x\_continuous(breaks = seq(min(total\_exports\_df$Year), max(total\_exports\_df$Year), by = 1)) +  
 guides(color = guide\_legend(override.aes = list(shape = c(16, 17), size = c(3, 4))),  
 linetype = guide\_legend(override.aes = list(size = 0.77)))+scale\_y\_continuous(breaks = seq(300, 500, by = 25))

## 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.

print(plot)



—————–SMART QUESTION 2—————————-

selected\_data <- exports\_data %>%  
 select(Country, F2018, F2019, F2020, F2021, F2022)  
head(exports\_data)

## ObjectId Country ISO2 Counterpart\_Country Counterpart\_ISO2  
## 1 3 Andorra, Principality of AD Belgium BE  
## 2 14 Andorra, Principality of AD France FR  
## 3 18 Andorra, Principality of AD Germany DE  
## 4 26 Andorra, Principality of AD Italy IT  
## 5 35 Andorra, Principality of AD Netherlands, The NL  
## 6 39 Andorra, Principality of AD Norway NO  
## Indicator Trade\_Flow F2012 F2013 F2014  
## 1 Environmental goods exports Exports 4977.802 17650.636 23345.22  
## 2 Environmental goods exports Exports 333741.685 654987.125 856655.32  
## 3 Environmental goods exports Exports 17916.401 16433.645 42109.43  
## 4 Environmental goods exports Exports 58748.693 195708.618 231083.88  
## 5 Environmental goods exports Exports 18278.694 9245.024 31706.86  
## 6 Environmental goods exports Exports 45672.751 40555.439 46440.88  
## F2015 F2016 F2017 F2018 F2019 F2020  
## 1 11296.40 19243.477 5714.549 12698.171 73595.050 1090.393  
## 2 1030334.07 1234084.840 1016697.537 1100321.851 1356900.374 2421820.712  
## 3 16199.17 36896.948 22514.621 1001.678 2921.677 12279.488  
## 4 86439.44 181035.927 483990.169 8248.304 2245.843 2856.204  
## 5 10940.51 8314.891 7467.299 2022.019 119551.294 27688.404  
## 6 15371.52 4304.702 5937.560 2805.334 2446.793 9370.652  
## F2021 F2022  
## 1 3052.548 347615.58  
## 2 2450100.158 2353566.87  
## 3 633214.245 142628.66  
## 4 7094.110 35013.62  
## 5 1176.319 23766.57  
## 6 86769.964 65582.46

head(selected\_data)

## Country F2018 F2019 F2020 F2021  
## 1 Andorra, Principality of 12698.171 73595.050 1090.393 3052.548  
## 2 Andorra, Principality of 1100321.851 1356900.374 2421820.712 2450100.158  
## 3 Andorra, Principality of 1001.678 2921.677 12279.488 633214.245  
## 4 Andorra, Principality of 8248.304 2245.843 2856.204 7094.110  
## 5 Andorra, Principality of 2022.019 119551.294 27688.404 1176.319  
## 6 Andorra, Principality of 2805.334 2446.793 9370.652 86769.964  
## F2022  
## 1 347615.58  
## 2 2353566.87  
## 3 142628.66  
## 4 35013.62  
## 5 23766.57  
## 6 65582.46

# Step 3: Check for each country if the exports have increased every year  
increasing\_exports <- selected\_data %>%  
 filter(F2018 < F2019 & F2019 < F2020 & F2020 < F2021 & F2021 < F2022) %>%  
 arrange(desc(F2022))

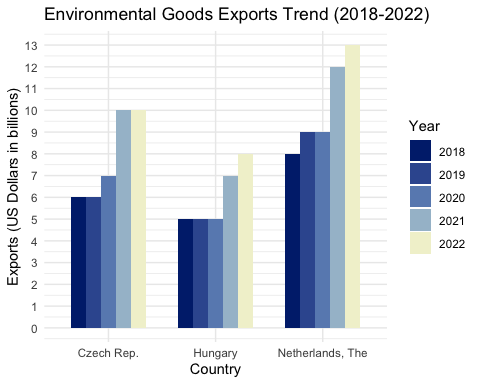
head(increasing\_exports)

## Country F2018 F2019 F2020 F2021 F2022  
## 1 Netherlands, The 8316603140 8772761559 9452270201 12181505503 13020268657  
## 2 Czech Rep. 6258020652 6390968996 7442721687 9633436831 10195826088  
## 3 Hungary 5162379027 5234097261 5457287764 6992357950 8399320651  
## 4 Netherlands, The 4112786985 5159306995 6431886896 7433899047 7556575458  
## 5 Belgium 2238178258 2663951782 3050020197 4039025943 4636423734  
## 6 Czech Rep. 1233755346 1281256479 1446627766 1996938588 2399384384

increasing\_exports\_long <- increasing\_exports %>%  
 gather(key = "Year", value = "Exports", -Country) %>%  
 mutate(Year = as.numeric(sub("F", "", Year)))   
  
increasing\_exports\_long$Exports=round(increasing\_exports\_long$Exports/1000000000)  
  
top\_three\_countries <- head(increasing\_exports, 3)  
  
  
top\_three\_countries\_names <- top\_three\_countries$Country  
  
top\_three\_countries\_long <- increasing\_exports\_long %>%  
 filter(Country %in% top\_three\_countries\_names)  
  
top\_three\_countries\_long <- top\_three\_countries\_long %>%  
 mutate(Country\_Year = paste(Country, Year, sep = "-"))  
head(top\_three\_countries\_long)

## Country Year Exports Country\_Year  
## 1 Netherlands, The 2018 8 Netherlands, The-2018  
## 2 Czech Rep. 2018 6 Czech Rep.-2018  
## 3 Hungary 2018 5 Hungary-2018  
## 4 Netherlands, The 2018 4 Netherlands, The-2018  
## 5 Czech Rep. 2018 1 Czech Rep.-2018  
## 6 Netherlands, The 2018 2 Netherlands, The-2018

ggplot(top\_three\_countries\_long, aes(x = Country, y = Exports, fill = as.factor(Year))) +  
 geom\_bar(stat = "identity", position = "dodge", width = 0.7) +  
 scale\_fill\_manual(values = c("2018" = "#00297b", "2019" = "#38599F", "2020" = "#698bbD", "2021" = "#A5bfd1", "2022" = "#f1f1d2")) +  
 labs(title = "Environmental Goods Exports Trend (2018-2022)",  
 x = "Country",  
 y = "Exports (US Dollars in billions)",  
 fill = "Year") +  
 theme\_minimal() +  
 scale\_y\_continuous(breaks = seq(0,15,by=1))



theme(legend.position = "bottom")

## List of 1  
## $ legend.position: chr "bottom"  
## - attr(\*, "class")= chr [1:2] "theme" "gg"  
## - attr(\*, "complete")= logi FALSE  
## - attr(\*, "validate")= logi TRUE

SMART QUESTION 04: How do landlocked countries compare to coastal countries in terms of environmental goods imports in 2022?

# Define the groups of countries  
landlocked\_countries\_group <- c('Czech Republic', 'Hungary', 'Azerbaijan', 'Bolivia', 'Georgia')  
coastal\_countries\_group <- c('Chile', 'Greece', 'Jordan', 'Estonia', 'Mozambique')  
  
# Filter data for the specified countries, indicators, and the year 2022  
filtered\_group\_data <- data %>%  
 filter(Country %in% c(landlocked\_countries\_group, coastal\_countries\_group),  
 Indicator %in% c('Environmental goods imports', 'Environmental goods exports'),  
 Trade\_Flow %in% c('Imports', 'Exports')) %>%  
 select(Country, Indicator, Trade\_Flow, F2022)  
  
# Display the filtered data  
head(filtered\_group\_data)

## Country Indicator Trade\_Flow F2022  
## 1 Bolivia Environmental goods imports Imports 97  
## 2 Bolivia Environmental goods exports Exports 2662444  
## 3 Bolivia Environmental goods imports Imports 27777260  
## 4 Bolivia Environmental goods imports Imports 337997  
## 5 Bolivia Environmental goods imports Imports 662011  
## 6 Bolivia Environmental goods imports Imports 130086

# Add a new column for country classification  
filtered\_group\_data$Country\_Classification <- ifelse(filtered\_group\_data$Country %in% landlocked\_countries\_group, 'Landlocked', 'Coastal')  
head(filtered\_group\_data)

## Country Indicator Trade\_Flow F2022  
## 1 Bolivia Environmental goods imports Imports 97  
## 2 Bolivia Environmental goods exports Exports 2662444  
## 3 Bolivia Environmental goods imports Imports 27777260  
## 4 Bolivia Environmental goods imports Imports 337997  
## 5 Bolivia Environmental goods imports Imports 662011  
## 6 Bolivia Environmental goods imports Imports 130086  
## Country\_Classification  
## 1 Landlocked  
## 2 Landlocked  
## 3 Landlocked  
## 4 Landlocked  
## 5 Landlocked  
## 6 Landlocked

# Calculate total trade (imports and exports) for landlocked and coastal groups  
total\_trade\_group <- filtered\_group\_data %>%  
 group\_by(Country\_Classification, Trade\_Flow) %>%  
 summarise(Total\_Trade = sum(F2022)) %>%  
 spread(key = Trade\_Flow, value = Total\_Trade)

## `summarise()` has grouped output by 'Country\_Classification'. You can override  
## using the `.groups` argument.

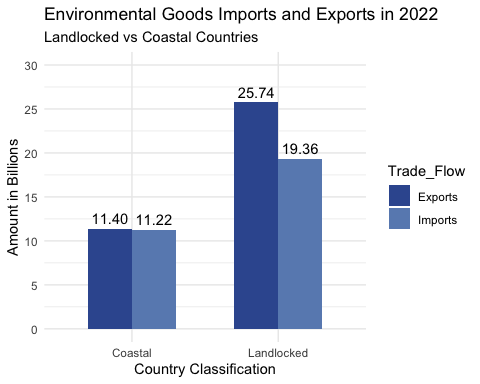
head(total\_trade\_group)

## # A tibble: 2 × 3  
## # Groups: Country\_Classification [2]  
## Country\_Classification Exports Imports  
## <chr> <dbl> <dbl>  
## 1 Coastal 11396243878. 11219542337.  
## 2 Landlocked 25736801925. 19357005298.

# Transform the data to a long format  
total\_trade\_long <- total\_trade\_group %>%  
 gather(key = "Trade\_Flow", value = "Amount", -Country\_Classification)  
head(total\_trade\_long)

## # A tibble: 4 × 3  
## # Groups: Country\_Classification [2]  
## Country\_Classification Trade\_Flow Amount  
## <chr> <chr> <dbl>  
## 1 Coastal Exports 11396243878.  
## 2 Landlocked Exports 25736801925.  
## 3 Coastal Imports 11219542337.  
## 4 Landlocked Imports 19357005298.

p <- ggplot(total\_trade\_long, aes(x = Country\_Classification, y = Amount, fill = Trade\_Flow)) +  
 geom\_col(position = "dodge", width = 0.6) +  
 scale\_y\_continuous(  
 labels = scales::label\_number(scale = 1e-9, accuracy = 1),  
 limits = c(0, 3e10),  
 breaks = seq(0, 3e10, by = 5e9)  
 ) +  
 labs(  
 y = "Amount in Billions",  
 x = "Country Classification",  
 title = "Environmental Goods Imports and Exports in 2022",  
 subtitle = "Landlocked vs Coastal Countries"  
 ) +  
 theme\_minimal() +  
 scale\_fill\_manual(values = c("Exports" = "#38599F", "Imports" = "#698bbd")) +  
 geom\_text(aes(label = scales::label\_number(accuracy = 0.01)(Amount/1e9)),  
 position = position\_dodge(width = 0.6), vjust = -0.5)  
  
# Show the plot  
p



——————–SMART QUESTION 05——————-

india\_trade <- filter(data, Counterpart\_Country == 'India')

environmental\_trade <- filter(india\_trade, Indicator %in% c('Environmental goods exports', 'Environmental goods imports'))

country\_indicator\_count <- environmental\_trade %>%  
 group\_by(Country) %>%  
 summarise(Unique\_Indicators = n\_distinct(Indicator))

countries\_with\_both <- filter(country\_indicator\_count, Unique\_Indicators == 2) %>%  
 pull(Country)

filtered\_data <- filter(environmental\_trade, Country %in% countries\_with\_both)  
  
# Step 2: Group the data by country and indicator, and calculate the sum for each group  
grouped\_data <- filtered\_data %>%  
 group\_by(Country, Indicator) %>%  
 summarise(Sum = sum(F2022, na.rm = TRUE)) %>%  
 ungroup()

## `summarise()` has grouped output by 'Country'. You can override using the  
## `.groups` argument.

# Step 3: Pivot the data to have exports and imports in separate columns for easier comparison  
trade\_sums <- grouped\_data %>%  
 spread(key = Indicator, value = Sum)  
  
# Step 4: Replace NaN values with 0  
trade\_sums[is.na(trade\_sums)] <- 0  
trade\_sums

## # A tibble: 32 × 3  
## Country `Environmental goods exports` Environmental goods im…¹  
## <chr> <dbl> <dbl>  
## 1 Armenia, Rep. of 108948. 2419733.  
## 2 Australia 66709661. 353861776.  
## 3 Belgium 176686852. 105309964.  
## 4 Bosnia and Herzegovina 227430. 10088742.  
## 5 Bulgaria 32707117. 5563577.  
## 6 Chile 286887. 110312620.  
## 7 China, P.R.: Hong Kong 984567750. 42102024.  
## 8 Croatia, Rep. of 10493427 20763636   
## 9 Cyprus 4028684. 4960816.  
## 10 Czech Rep. 206991050 130838923   
## # ℹ 22 more rows  
## # ℹ abbreviated name: ¹​`Environmental goods imports`

#Step 1: Calculate the Import-Export Difference  
trade\_sums <- trade\_sums %>%  
 mutate(Import\_Export\_Difference = `Environmental goods imports` - `Environmental goods exports`)  
  
# Step 2: Select Specific Columns  
result <- select(trade\_sums, Country, Import\_Export\_Difference)  
head(result)

## # A tibble: 6 × 2  
## Country Import\_Export\_Difference  
## <chr> <dbl>  
## 1 Armenia, Rep. of 2310784.  
## 2 Australia 287152114.  
## 3 Belgium -71376887.  
## 4 Bosnia and Herzegovina 9861312.  
## 5 Bulgaria -27143540.  
## 6 Chile 110025733.

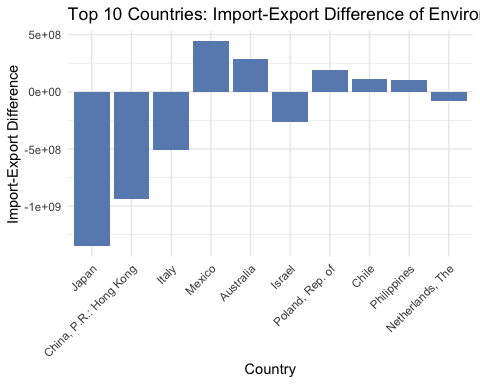
trade\_values <- trade\_sums %>%  
 pivot\_longer(cols = c(`Environmental goods exports`, `Environmental goods imports`),  
 names\_to = "Indicator",  
 values\_to = "Trade\_Value")  
  
# Showing the trade values  
print(trade\_values)

## # A tibble: 64 × 4  
## Country Import\_Export\_Difference Indicator Trade\_Value  
## <chr> <dbl> <chr> <dbl>  
## 1 Armenia, Rep. of 2310784. Environmental go… 108948.  
## 2 Armenia, Rep. of 2310784. Environmental go… 2419733.  
## 3 Australia 287152114. Environmental go… 66709661.  
## 4 Australia 287152114. Environmental go… 353861776.  
## 5 Belgium -71376887. Environmental go… 176686852.  
## 6 Belgium -71376887. Environmental go… 105309964.  
## 7 Bosnia and Herzegovina 9861312. Environmental go… 227430.  
## 8 Bosnia and Herzegovina 9861312. Environmental go… 10088742.  
## 9 Bulgaria -27143540. Environmental go… 32707117.  
## 10 Bulgaria -27143540. Environmental go… 5563577.  
## # ℹ 54 more rows

# Step 1: Calculate the Import-Export Difference  
trade\_sums <- trade\_sums %>%  
 mutate(Import\_Export\_Difference = `Environmental goods imports` - `Environmental goods exports`)  
  
# Step 2: Select Specific Columns  
result <- select(trade\_sums, Country, Import\_Export\_Difference)  
head(result)

## # A tibble: 6 × 2  
## Country Import\_Export\_Difference  
## <chr> <dbl>  
## 1 Armenia, Rep. of 2310784.  
## 2 Australia 287152114.  
## 3 Belgium -71376887.  
## 4 Bosnia and Herzegovina 9861312.  
## 5 Bulgaria -27143540.  
## 6 Chile 110025733.

library(ggplot2)  
  
top\_countries <- result %>%  
 arrange(desc(abs(Import\_Export\_Difference)))  
  
# Step 2: Select the Top 10  
top\_10\_countries <- head(top\_countries, 10)  
  
# Step 3: Create the Plot  
ggplot(data = top\_10\_countries, aes(x = reorder(Country, -abs(Import\_Export\_Difference)), y = Import\_Export\_Difference)) +  
 geom\_bar(stat = "identity", fill = "#698bbd") +  
 labs(title = "Top 10 Countries: Import-Export Difference of Environmental Goods",  
 x = "Country",  
 y = "Import-Export Difference") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



— SMART QUESTION 06

trade\_balance\_data <- filter(data, data$Indicator == 'Environmental goods trade balance')  
head(trade\_balance\_data)

## ObjectId Country ISO2 Counterpart\_Country Counterpart\_ISO2  
## 1 5 Andorra, Principality of AD Belgium BE  
## 2 16 Andorra, Principality of AD France FR  
## 3 20 Andorra, Principality of AD Germany DE  
## 4 28 Andorra, Principality of AD Italy IT  
## 5 37 Andorra, Principality of AD Netherlands, The NL  
## 6 45 Andorra, Principality of AD Portugal PT  
## Indicator Trade\_Flow F2012 F2013  
## 1 Environmental goods trade balance Not Applicable -50272.39 -46415.45  
## 2 Environmental goods trade balance Not Applicable -4660354.05 -3213472.14  
## 3 Environmental goods trade balance Not Applicable -902931.22 -1180312.12  
## 4 Environmental goods trade balance Not Applicable -261423.82 -1476528.65  
## 5 Environmental goods trade balance Not Applicable -111084.27 -99883.18  
## 6 Environmental goods trade balance Not Applicable 23861.84 15222.90  
## F2014 F2015 F2016 F2017 F2018 F2019  
## 1 -116914.44 -40948.884 -105166.29 -61199.687 -68061.058 9569.782  
## 2 -4016035.12 -3518875.064 -2807624.24 -3694204.514 -6275971.269 -3616339.936  
## 3 -4899167.99 -1255385.468 -770787.72 -1270529.379 -1347615.072 -1140457.669  
## 4 -207582.49 -231325.854 -738532.34 -319882.415 -807337.711 -368964.385  
## 5 -68016.72 -30976.790 -55366.03 -573365.953 -651176.316 -469429.290  
## 6 28673.39 8721.299 18083.88 7228.336 1390.928 -19719.412  
## F2020 F2021 F2022  
## 1 -44764.41 -165124.2 174430.8  
## 2 -1322269.04 -1923390.6 -6057906.2  
## 3 -2892163.58 -6314674.4 -12228802.0  
## 4 -1109829.47 -2885661.9 -1856007.4  
## 5 -324642.07 -179536.8 -407419.8  
## 6 -25041.73 -424260.7 -123822.2

asian\_countries = c('Afghanistan', 'Armenia', 'Azerbaijan', 'Bahrain', 'Bangladesh', 'Bhutan', 'Brunei Darussalam', 'Cambodia', 'China', 'Cyprus', 'Georgia', 'India', 'Indonesia', 'Iran (Islamic Republic of)', 'Iraq', 'Israel', 'Japan', 'Jordan', 'Kazakhstan', 'Kuwait', 'Kyrgyzstan', 'Lao People\'s Democratic Republic', 'Lebanon', 'Malaysia', 'Maldives', 'Mongolia', 'Myanmar', 'Nepal', 'Korea (Democratic People\'s Republic of)', 'Oman', 'Pakistan', 'Palestine', 'Philippines', 'Qatar', 'Saudi Arabia', 'Singapore', 'Korea (Republic of)', 'Sri Lanka', 'Syrian Arab Republic', 'Taiwan', 'Tajikistan', 'Thailand', 'Timor-Leste', 'Turkey', 'Turkmenistan', 'United Arab Emirates', 'Uzbekistan', 'Viet Nam', 'Yemen')  
  
european\_countries = c('Albania', 'Andorra', 'Austria', 'Belarus', 'Belgium', 'Bosnia and Herzegovina', 'Bulgaria', 'Croatia', 'Cyprus', 'Czech Republic', 'Denmark', 'Estonia', 'Finland', 'France', 'Germany', 'Greece', 'Hungary', 'Iceland', 'Ireland', 'Italy', 'Latvia', 'Liechtenstein', 'Lithuania', 'Luxembourg', 'Malta', 'Moldova', 'Monaco', 'Montenegro', 'Netherlands', 'North Macedonia', 'Norway', 'Poland', 'Portugal', 'Romania', 'Russian Federation', 'San Marino', 'Serbia', 'Slovakia', 'Slovenia', 'Spain', 'Sweden', 'Switzerland', 'Ukraine', 'United Kingdom', 'Vatican City State (Holy See)')

trade\_balance\_data$Region <- ifelse(trade\_balance\_data$Country %in% asian\_countries, 'Asia',  
 ifelse(trade\_balance\_data$Country %in% european\_countries, 'Europe', 'Other'))  
head(trade\_balance\_data)

## ObjectId Country ISO2 Counterpart\_Country Counterpart\_ISO2  
## 1 5 Andorra, Principality of AD Belgium BE  
## 2 16 Andorra, Principality of AD France FR  
## 3 20 Andorra, Principality of AD Germany DE  
## 4 28 Andorra, Principality of AD Italy IT  
## 5 37 Andorra, Principality of AD Netherlands, The NL  
## 6 45 Andorra, Principality of AD Portugal PT  
## Indicator Trade\_Flow F2012 F2013  
## 1 Environmental goods trade balance Not Applicable -50272.39 -46415.45  
## 2 Environmental goods trade balance Not Applicable -4660354.05 -3213472.14  
## 3 Environmental goods trade balance Not Applicable -902931.22 -1180312.12  
## 4 Environmental goods trade balance Not Applicable -261423.82 -1476528.65  
## 5 Environmental goods trade balance Not Applicable -111084.27 -99883.18  
## 6 Environmental goods trade balance Not Applicable 23861.84 15222.90  
## F2014 F2015 F2016 F2017 F2018 F2019  
## 1 -116914.44 -40948.884 -105166.29 -61199.687 -68061.058 9569.782  
## 2 -4016035.12 -3518875.064 -2807624.24 -3694204.514 -6275971.269 -3616339.936  
## 3 -4899167.99 -1255385.468 -770787.72 -1270529.379 -1347615.072 -1140457.669  
## 4 -207582.49 -231325.854 -738532.34 -319882.415 -807337.711 -368964.385  
## 5 -68016.72 -30976.790 -55366.03 -573365.953 -651176.316 -469429.290  
## 6 28673.39 8721.299 18083.88 7228.336 1390.928 -19719.412  
## F2020 F2021 F2022 Region  
## 1 -44764.41 -165124.2 174430.8 Other  
## 2 -1322269.04 -1923390.6 -6057906.2 Other  
## 3 -2892163.58 -6314674.4 -12228802.0 Other  
## 4 -1109829.47 -2885661.9 -1856007.4 Other  
## 5 -324642.07 -179536.8 -407419.8 Other  
## 6 -25041.73 -424260.7 -123822.2 Other

# Select the columns for years and Region  
years <- c('F2015', 'F2016', 'F2017', 'F2018', 'F2019', 'F2020', 'F2021', 'F2022')  
agg\_data <- trade\_balance\_data %>%  
 select(one\_of(c("Region", years)))  
# Group by 'Region' and calculate the sum for each year  
agg\_data <- agg\_data %>%  
 group\_by(Region) %>%  
 summarise(across(.cols = everything(), .fns = sum, na.rm = TRUE))

## Warning: There was 1 warning in `summarise()`.  
## ℹ In argument: `across(.cols = everything(), .fns = sum, na.rm = TRUE)`.  
## ℹ In group 1: `Region = "Asia"`.  
## Caused by warning:  
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.  
## Supply arguments directly to `.fns` through an anonymous function instead.  
##   
## # Previously  
## across(a:b, mean, na.rm = TRUE)  
##   
## # Now  
## across(a:b, \(x) mean(x, na.rm = TRUE))

# Filter only 'Asia' and 'Europe'  
agg\_data <- agg\_data %>%  
 filter(Region %in% c('Asia', 'Europe'))  
head(agg\_data)

## # A tibble: 2 × 9  
## Region F2015 F2016 F2017 F2018 F2019 F2020 F2021 F2022  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Asia 32196601464. 3.48e10 5.59e10 5.60e10 4.85e10 4.94e10 5.72e10 4.35e10  
## 2 Europe 26128055650. 2.56e10 1.94e10 1.58e10 1.51e10 7.07e 9 1.88e 7 -7.17e 9

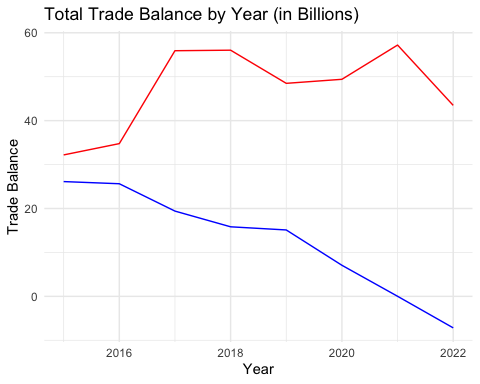
result6 <- data.frame(Year = 2015:2022,   
 'Asia' = t(agg\_data[1,2:ncol(agg\_data)]),   
 'Europe' = t(agg\_data[2,2:ncol(agg\_data)]))  
head(result6)

## Year Asia Europe  
## F2015 2015 32196601464 26128055650  
## F2016 2016 34770140193 25635947569  
## F2017 2017 55917580807 19411082983  
## F2018 2018 56041252955 15834022373  
## F2019 2019 48492246199 15115300322  
## F2020 2020 49410870490 7074862160

ggplot(result6, aes(x = Year)) +  
 geom\_line(aes(y = Asia / 1000000000, label='Asia'), color = "red", linetype = "solid") +  
 geom\_line(aes(y = Europe / 1000000000, label='Europe'), color = "blue", linetype = "solid") +  
 labs(  
 title = "Total Trade Balance by Year (in Billions)",  
 x = "Year",  
 y = "Trade Balance"  
 ) +  
 scale\_color\_manual(values = c("Asia" = "red", "Europe" = "blue"),  
 name = "Legend",  
 labels = c('Asia', 'Europe')) +  
 theme\_minimal()

## Warning in geom\_line(aes(y = Asia/1e+09, label = "Asia"), color = "red", :  
## Ignoring unknown aesthetics: label

## Warning in geom\_line(aes(y = Europe/1e+09, label = "Europe"), color = "blue", :  
## Ignoring unknown aesthetics: label



correlation\_coefficient <- cor(result6['Asia'], result6['Europe'])  
cat("Pearson Correlation Coefficient:", correlation\_coefficient, "\n")

## Pearson Correlation Coefficient: -0.4190462

— SMART QUESTION 07

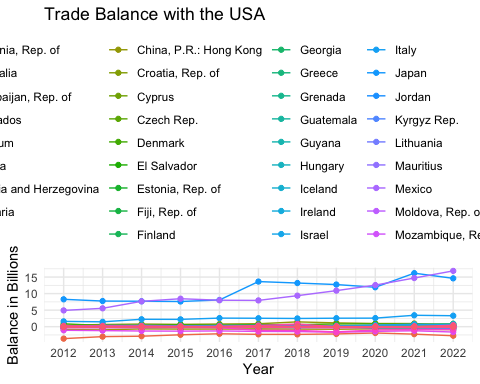
us\_data = filter(trade\_balance\_data, Counterpart\_Country == "United States")  
us\_data <- us\_data %>%   
 select("Country", "F2012", "F2013", "F2014", "F2015", "F2016", "F2017", "F2018", "F2019", "F2020", "F2021", "F2022")  
head(us\_data)

## Country F2012 F2013 F2014 F2015  
## 1 Armenia, Rep. of -7149680 -6718457 -13780041 -6287212  
## 2 Australia -3611066455 -3005607704 -2877301532 -2448904759  
## 3 Azerbaijan, Rep. of -92008383 -64933265 -215700032 -157465963  
## 4 Barbados -34432082 -35730795 -38109436 -35910046  
## 5 Belgium -265501695 -187855979 -184250438 -187937311  
## 6 Bolivia -182974014 -219653681 -185938002 -192285367  
## F2016 F2017 F2018 F2019 F2020 F2021  
## 1 -4273964 -7148792 -5344556 -5066965 -4878982 -3967025  
## 2 -2142689380 -2314985441 -2325780846 -2195466968 -1902978695 -2256888884  
## 3 -185216637 -88568664 -104838339 -73901788 -75768338 -82637715  
## 4 -37215323 -39844124 -36964878 -37917177 -43993174 -53083028  
## 5 -197501805 -355605586 -168434922 -136639816 -292272712 -144103700  
## 6 -113046672 -105158665 -102837114 -107993682 -59337147 -59301829  
## F2022  
## 1 -22651737  
## 2 -2735994886  
## 3 -120225630  
## 4 -55596984  
## 5 689975026  
## 6 -62693025

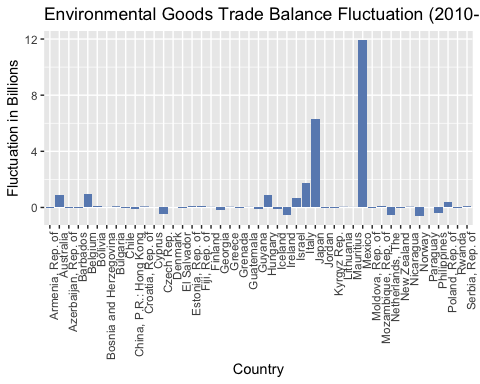
# Convert the data into a long format  
trade\_balance\_data\_long <- pivot\_longer(us\_data,   
 cols = starts\_with("F"),   
 names\_to = "Year",   
 values\_to = "Balance")  
trade\_balance\_data\_long$Year <- as.numeric(substring(trade\_balance\_data\_long$Year, 2))  
head(trade\_balance\_data\_long)

## # A tibble: 6 × 3  
## Country Year Balance  
## <chr> <dbl> <dbl>  
## 1 Armenia, Rep. of 2012 -7149680   
## 2 Armenia, Rep. of 2013 -6718457   
## 3 Armenia, Rep. of 2014 -13780041   
## 4 Armenia, Rep. of 2015 -6287212   
## 5 Armenia, Rep. of 2016 -4273964   
## 6 Armenia, Rep. of 2017 -7148792.

# Create the multiline graph  
ggplot(trade\_balance\_data\_long, aes(x = Year, y = Balance / 1000000000, color = Country)) +  
 geom\_line() +   
 geom\_point() +  
 labs(  
 title = "Trade Balance with the USA",  
 x = "Year",  
 y = "Balance in Billions"  
 ) +  
 theme\_minimal() +  
 theme(legend.position = "top") +  
 scale\_x\_continuous(breaks = seq(min(as.numeric(trade\_balance\_data\_long$Year)),   
 max(as.numeric(trade\_balance\_data\_long$Year)), 1))



fluctuation\_data <- us\_data %>%  
 mutate(Fluctuation = F2022 - F2012)  
  
ggplot(fluctuation\_data, aes(x = Country, y = Fluctuation / 1000000000)) +  
 geom\_bar(stat = "identity", fill = "#698bbd") +  
 labs(  
 title = "Environmental Goods Trade Balance Fluctuation (2010-2022)",  
 x = "Country",  
 y = "Fluctuation in Billions"  
 ) +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))



max\_fluctuation\_index <- which.max(abs(fluctuation\_data$Fluctuation))  
country\_with\_max\_fluctuation <- fluctuation\_data$Country[max\_fluctuation\_index]  
max\_fluctuation <- fluctuation\_data$Fluctuation[max\_fluctuation\_index]  
  
cat("Country with Maximum Fluctuation (based on absolute values):", country\_with\_max\_fluctuation, "\n")

## Country with Maximum Fluctuation (based on absolute values): Mexico

cat("Maximum Fluctuation Value (based on absolute values in billions):", max\_fluctuation / 1000000000, "\n")

## Maximum Fluctuation Value (based on absolute values in billions): 11.94997

——————SMART QUESTION 08—————————-

#filtered\_data = data[['Country', 'Counterpart\_Country', 'Trade\_Flow', 'F2022']]  
filtered\_data <- data %>%  
 select(Country, Counterpart\_Country, Trade\_Flow, F2022)  
head(filtered\_data)

## Country Counterpart\_Country Trade\_Flow F2022  
## 1 Andorra, Principality of Andorra, Principality of Imports 12337.32  
## 2 Andorra, Principality of Austria Imports 311349.64  
## 3 Andorra, Principality of Belgium Exports 347615.58  
## 4 Andorra, Principality of Belgium Imports 173184.81  
## 5 Andorra, Principality of Belgium Not Applicable 174430.77  
## 6 Andorra, Principality of Belgium Not Applicable 520800.40

#calculate mutual trade volume.  
filtered\_data\_2 <- filtered\_data %>%  
 mutate(Country\_Pair = if\_else(Country < Counterpart\_Country,  
 paste(Country, Counterpart\_Country, sep = " - "),  
 paste(Counterpart\_Country, Country, sep = " - ")))  
head(filtered\_data\_2)

## Country Counterpart\_Country Trade\_Flow F2022  
## 1 Andorra, Principality of Andorra, Principality of Imports 12337.32  
## 2 Andorra, Principality of Austria Imports 311349.64  
## 3 Andorra, Principality of Belgium Exports 347615.58  
## 4 Andorra, Principality of Belgium Imports 173184.81  
## 5 Andorra, Principality of Belgium Not Applicable 174430.77  
## 6 Andorra, Principality of Belgium Not Applicable 520800.40  
## Country\_Pair  
## 1 Andorra, Principality of - Andorra, Principality of  
## 2 Andorra, Principality of - Austria  
## 3 Andorra, Principality of - Belgium  
## 4 Andorra, Principality of - Belgium  
## 5 Andorra, Principality of - Belgium  
## 6 Andorra, Principality of - Belgium

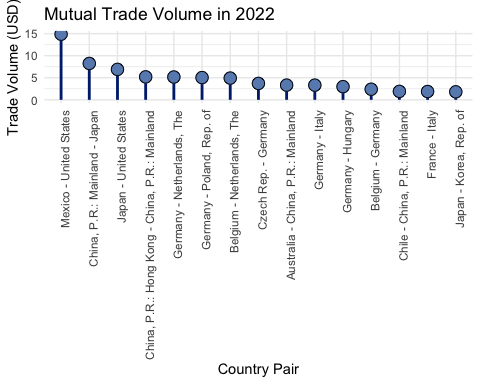
# Calculate the mutual trade volume for each country pair  
mutual\_trade\_volume <- filtered\_data\_2 %>%  
 group\_by(Country\_Pair) %>%  
 summarise(Total\_Trade\_2022 = sum(F2022, na.rm = TRUE)) %>%  
 ungroup()  
head(mutual\_trade\_volume)

## # A tibble: 6 × 2  
## Country\_Pair Total\_Trade\_2022  
## <chr> <dbl>  
## 1 Afghanistan, Islamic Rep. of - Australia 44139.  
## 2 Afghanistan, Islamic Rep. of - Belgium 56133.  
## 3 Afghanistan, Islamic Rep. of - Bulgaria 946.  
## 4 Afghanistan, Islamic Rep. of - China, P.R.: Hong Kong 6097.  
## 5 Afghanistan, Islamic Rep. of - Czech Rep. 2878.  
## 6 Afghanistan, Islamic Rep. of - Denmark 153246.

#sort the results and find the country pairs with the highest mutual trade volume in 2022  
highest\_mutual\_trade\_volume <- mutual\_trade\_volume %>%  
 arrange(desc(Total\_Trade\_2022))  
highest\_mutual\_trade\_volume<- head(highest\_mutual\_trade\_volume,15)  
highest\_mutual\_trade\_volume$Total\_Trade\_2022<- highest\_mutual\_trade\_volume$Total\_Trade\_2022/10000000000  
head(highest\_mutual\_trade\_volume)

## # A tibble: 6 × 2  
## Country\_Pair Total\_Trade\_2022  
## <chr> <dbl>  
## 1 Mexico - United States 14.9   
## 2 China, P.R.: Mainland - Japan 8.24  
## 3 Japan - United States 6.90  
## 4 China, P.R.: Hong Kong - China, P.R.: Mainland 5.20  
## 5 Germany - Netherlands, The 5.17  
## 6 Germany - Poland, Rep. of 5.04

ggplot(highest\_mutual\_trade\_volume, aes(x = reorder(Country\_Pair, -Total\_Trade\_2022), y = Total\_Trade\_2022)) +  
 geom\_segment(aes(x = reorder(Country\_Pair, -Total\_Trade\_2022), xend = reorder(Country\_Pair, -Total\_Trade\_2022), y = 0, yend = Total\_Trade\_2022), color = "#00297b", size = 1) +  
 geom\_point(aes(x = reorder(Country\_Pair, -Total\_Trade\_2022), y = Total\_Trade\_2022), color = "#698bbd", size = 4) +  
 geom\_point(aes(x = reorder(Country\_Pair, -Total\_Trade\_2022), y = Total\_Trade\_2022), color = "black", size = 4, shape = 1) +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 1, hjust=1)) +  
 labs(title = "Mutual Trade Volume in 2022",  
 x = "Country Pair",  
 y = "Trade Volume (USD)")



Statistical Testing

Hypothesis: Is there a significant difference in mutual trade volumes across different country-counterpart pairs in 2022?”

Test: ANNOVA # Step 1: Prepare the data for ANOVA

dataset <- data %>%  
 group\_by(Country, Counterpart\_Country) %>%  
 mutate(Mutual\_Trade\_2022 = sum(F2022, na.rm = TRUE)) %>%  
 ungroup()

unique\_pairs <- dataset %>%  
 distinct(Country, Counterpart\_Country, .keep\_all = TRUE)  
sorted\_pairs <- unique\_pairs %>%  
 arrange(desc(Mutual\_Trade\_2022))  
top\_pairs <- head(sorted\_pairs, 15)  
head(top\_pairs)

## # A tibble: 6 × 19  
## ObjectId Country ISO2 Counterpart\_Country Counterpart\_ISO2 Indicator  
## <int> <chr> <chr> <chr> <chr> <chr>   
## 1 14776 Mexico MX United States US Environm…  
## 2 12644 Japan JP China, P.R.: Mainl… CN Environm…  
## 3 13185 Japan JP United States US Environm…  
## 4 3852 China, P.R.: Ho… HK China, P.R.: Mainl… CN Environm…  
## 5 15717 Netherlands, The NL Germany DE Environm…  
## 6 18640 Poland, Rep. of PL Germany DE Environm…  
## # ℹ 13 more variables: Trade\_Flow <chr>, F2012 <dbl>, F2013 <dbl>, F2014 <dbl>,  
## # F2015 <dbl>, F2016 <dbl>, F2017 <dbl>, F2018 <dbl>, F2019 <dbl>,  
## # F2020 <dbl>, F2021 <dbl>, F2022 <dbl>, Mutual\_Trade\_2022 <dbl>

anova\_data <- lapply(1:nrow(top\_pairs), function(i) {  
 row <- top\_pairs[i, ]  
 subset\_data <- subset(dataset, Country == row$Country & Counterpart\_Country == row$Counterpart\_Country)  
 c(subset\_data$F2012, subset\_data$F2013, subset\_data$F2014, subset\_data$F2015, subset\_data$F2016,  
 subset\_data$F2017, subset\_data$F2018, subset\_data$F2019, subset\_data$F2020, subset\_data$F2021, subset\_data$F2022)  
})

anova\_df <- data.frame(  
 values = unlist(anova\_data),  
 group = factor(rep(1:10, each = 11))  
)  
head(anova\_df)

## values group  
## 1 2.388243e+10 1  
## 2 8.297059e+00 1  
## 3 1.896264e+10 1  
## 4 9.664146e+00 1  
## 5 4.919792e+09 1  
## 6 4.284507e+10 1

# Performing the ANOVA test  
anova\_result <- aov(values ~ group, data = anova\_df)  
summary(anova\_result)

## Df Sum Sq Mean Sq F value Pr(>F)  
## group 9 5.179e+20 5.755e+19 0.621 0.78  
## Residuals 980 9.078e+22 9.263e+19

Conclusion: Because the p-value (0.78) is much greater than the common alpha level of 0.05, we fail to reject the null hypothesis. There is not enough statistical evidence to conclude that there is a significant difference in mutual trade volumes across different country-counterpart pairs in 2022. Interpretation: The mutual trade volumes across the selected country-counterpart pairs in 2022 do not significantly differ from each other, based on the ANOVA test conducted on the top trading pairs. This suggests that the variations in trade volumes within the pairs are not significantly less than the variations between the pairs.

——————————-SMART QUESTION 9——————————————————

#filtered\_data = data[['Country', 'Counterpart\_Country', 'Trade\_Flow', 'F2022']]  
filtered\_data <- data %>%  
 select(Country, Counterpart\_Country, Trade\_Flow, F2022)  
head(filtered\_data)

## Country Counterpart\_Country Trade\_Flow F2022  
## 1 Andorra, Principality of Andorra, Principality of Imports 12337.32  
## 2 Andorra, Principality of Austria Imports 311349.64  
## 3 Andorra, Principality of Belgium Exports 347615.58  
## 4 Andorra, Principality of Belgium Imports 173184.81  
## 5 Andorra, Principality of Belgium Not Applicable 174430.77  
## 6 Andorra, Principality of Belgium Not Applicable 520800.40

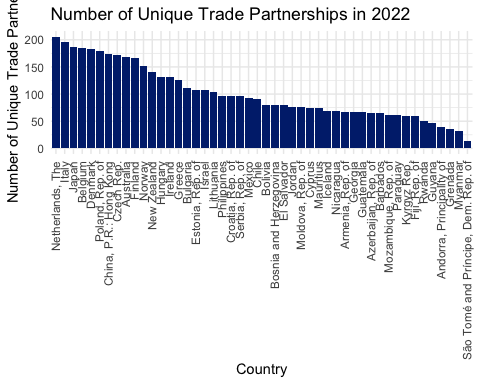
trade\_partnerships <- filtered\_data %>%  
 group\_by(Country) %>%  
 summarise(Unique\_Trade\_Partners = n\_distinct(Counterpart\_Country)) %>%  
 ungroup()  
head(trade\_partnerships)

## # A tibble: 6 × 2  
## Country Unique\_Trade\_Partners  
## <chr> <int>  
## 1 Andorra, Principality of 39  
## 2 Armenia, Rep. of 67  
## 3 Australia 169  
## 4 Azerbaijan, Rep. of 66  
## 5 Barbados 65  
## 6 Belgium 185

#Find the Country with the Most Diversified Trade  
  
most\_diversified\_trade = most\_diversified\_trade <- trade\_partnerships %>%  
 arrange(desc(Unique\_Trade\_Partners))  
head(most\_diversified\_trade)

## # A tibble: 6 × 2  
## Country Unique\_Trade\_Partners  
## <chr> <int>  
## 1 Netherlands, The 206  
## 2 Italy 195  
## 3 Japan 187  
## 4 Belgium 185  
## 5 Denmark 183  
## 6 Poland, Rep. of 180

# Bar Chart  
  
  
# Bar Chart  
ggplot(most\_diversified\_trade, aes(x = reorder(Country, -Unique\_Trade\_Partners), y = Unique\_Trade\_Partners)) +  
 geom\_bar(stat = "identity", fill = "#00297b") +  
 theme\_minimal() +  
 labs(title = "Number of Unique Trade Partnerships in 2022",  
 x = "Country",  
 y = "Number of Unique Trade Partners") +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))

 > Statistical Testing

Herfindahl-Hirschman Index (HHI):

head(most\_diversified\_trade)

## # A tibble: 6 × 2  
## Country Unique\_Trade\_Partners  
## <chr> <int>  
## 1 Netherlands, The 206  
## 2 Italy 195  
## 3 Japan 187  
## 4 Belgium 185  
## 5 Denmark 183  
## 6 Poland, Rep. of 180

total\_trade <- data %>%  
 group\_by(Country) %>%  
 summarise(Total\_Trade\_2022 = sum(F2022, na.rm = TRUE))  
head(total\_trade)

## # A tibble: 6 × 2  
## Country Total\_Trade\_2022  
## <chr> <dbl>  
## 1 Andorra, Principality of 106828803.  
## 2 Armenia, Rep. of 601531474.  
## 3 Australia 61311121447.  
## 4 Azerbaijan, Rep. of 1420948623.  
## 5 Barbados 204736521.  
## 6 Belgium 135814661809.

trade\_proportions <- data %>%  
 left\_join(total\_trade, by = "Country") %>%  
 mutate(Trade\_Proportion = F2022 / Total\_Trade\_2022)  
head(trade\_proportions)

## ObjectId Country ISO2 Counterpart\_Country  
## 1 1 Andorra, Principality of AD Andorra, Principality of  
## 2 2 Andorra, Principality of AD Austria  
## 3 3 Andorra, Principality of AD Belgium  
## 4 4 Andorra, Principality of AD Belgium  
## 5 5 Andorra, Principality of AD Belgium  
## 6 6 Andorra, Principality of AD Belgium  
## Counterpart\_ISO2 Indicator Trade\_Flow F2012  
## 1 AD Environmental goods imports Imports 124176.704  
## 2 AT Environmental goods imports Imports 57436.525  
## 3 BE Environmental goods exports Exports 4977.802  
## 4 BE Environmental goods imports Imports 55250.193  
## 5 BE Environmental goods trade balance Not Applicable -50272.391  
## 6 BE Total trade in environmental goods Not Applicable 60227.995  
## F2013 F2014 F2015 F2016 F2017 F2018 F2019  
## 1 72869.57 15552.94 688.099 408.981 8042.192 4921.683 831.419  
## 2 79987.58 616090.09 69688.266 123826.566 266202.777 646338.649 471439.576  
## 3 17650.64 23345.22 11296.396 19243.477 5714.549 12698.171 73595.050  
## 4 64066.09 140259.66 52245.280 124409.763 66914.236 80759.229 64025.268  
## 5 -46415.45 -116914.44 -40948.884 -105166.286 -61199.687 -68061.058 9569.782  
## 6 81716.73 163604.87 63541.676 143653.240 72628.785 93457.400 137620.318  
## F2020 F2021 F2022 Total\_Trade\_2022 Trade\_Proportion  
## 1 9112.604 6622.224 12337.32 106828803 0.0001154868  
## 2 456033.149 1264195.860 311349.64 106828803 0.0029144728  
## 3 1090.393 3052.548 347615.58 106828803 0.0032539500  
## 4 45854.805 168176.728 173184.81 106828803 0.0016211434  
## 5 -44764.412 -165124.180 174430.77 106828803 0.0016328065  
## 6 46945.198 171229.276 520800.40 106828803 0.0048750934

# Assuming trade\_proportions is your data frame and is already available in your R environment  
hhi\_results <- trade\_proportions %>%  
 group\_by(Country) %>%  
 summarise(HHI = round(sum(Trade\_Proportion^2), 2))  
  
# Now hhi\_results will contain the HHI values rounded to two decimal places  
print(hhi\_results)

## # A tibble: 48 × 2  
## Country HHI  
## <chr> <dbl>  
## 1 Andorra, Principality of 0.82  
## 2 Armenia, Rep. of 0.13  
## 3 Australia 0.13  
## 4 Azerbaijan, Rep. of 0.12  
## 5 Barbados 0.32  
## 6 Belgium 0.03  
## 7 Bolivia 0.17  
## 8 Bosnia and Herzegovina 0.05  
## 9 Bulgaria 0.03  
## 10 Chile 0.11  
## # ℹ 38 more rows

hhi\_data <- hhi\_results %>%  
 mutate(Group = case\_when(  
 hhi\_results$HHI >= 0.25 ~ "Low Diversity",  
 hhi\_results$HHI >= 0.15 & hhi\_results$HHI < 0.25 ~ "Moderate Diversity",  
 hhi\_results$HHI < 0.15 ~ "High Diversity"  
 ))  
head(hhi\_data)

## # A tibble: 6 × 3  
## Country HHI Group   
## <chr> <dbl> <chr>   
## 1 Andorra, Principality of 0.82 Low Diversity   
## 2 Armenia, Rep. of 0.13 High Diversity  
## 3 Australia 0.13 High Diversity  
## 4 Azerbaijan, Rep. of 0.12 High Diversity  
## 5 Barbados 0.32 Low Diversity   
## 6 Belgium 0.03 High Diversity

hhi\_data$Group <- as.factor(hhi\_data$Group)  
head(hhi\_data)

## # A tibble: 6 × 3  
## Country HHI Group   
## <chr> <dbl> <fct>   
## 1 Andorra, Principality of 0.82 Low Diversity   
## 2 Armenia, Rep. of 0.13 High Diversity  
## 3 Australia 0.13 High Diversity  
## 4 Azerbaijan, Rep. of 0.12 High Diversity  
## 5 Barbados 0.32 Low Diversity   
## 6 Belgium 0.03 High Diversity

anova\_result <- aov(HHI ~ Group, data = hhi\_data)  
summary(anova\_result)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Group 2 1.2363 0.6182 71.75 1.01e-14 \*\*\*  
## Residuals 45 0.3877 0.0086   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Null Hypothesis (H0): All countries have the same level of diversity in their environmental goods trade partnerships in 2022.

Alternative Hypothesis (H1): There is at least one country that has a significantly different level of diversity in its environmental goods trade partnerships in 2022 compared to other countries.

Conclusion: This means that there is a statistically significant difference in the diversity of environmental goods trade partnerships among the countries in 2022.