R Coding Sample

Jonas Wallstein

2022-11-20

The goal of this code is to find similiarities in emission profiles of several countries by performing a Principal Component Analysis on emission, GDP and climate pledge data

```
# Setup
library(readxl)
library(reshape)

data_path = paste(dirname(getwd()), "/data/", sep = "")
```

```
filename_emissions = paste(data_path, "climatetrace_emissions_by_subsector_timeseries_interval_year_sin

emissions = read.csv(filename_emissions,sep = ",", header = TRUE)

# Aggregating Emissions over each Country
total_emissions = aggregate(emissions[,1], by = list(emissions$country, emissions$country_full), FUN = s
# Renaming variables
colnames(total_emissions) = c("Country Code","Emissions Data Country Name", "Emissions")

# Emissions per sector
per_sector_emissions = aggregate(emissions[,1], by = list(emissions$country, emissions$country_full, em
colnames(per_sector_emissions) = c("country code","country name", "sector", "emissions")
per_sector_emissions = reshape(per_sector_emissions, idvar = c("country code", "country name") , timeva
per_sector_emissions$emissions.total = total_emissions$Emissions


head(per_sector_emissions)
```

1. Create a dataset on sector-level emissions from data from the climate trace project: https://climatetrace.org/

```
##   country code   country name emissions.agriculture emissions.buildings
## 1          AFG    Afghanistan              15406949              714200
## 2          ALA  Åland Islands                     0                   0
## 3          ALB        Albania               2899067              723815
## 4          DZA        Algeria              11578455            27249658
## 5          ASM American Samoa                     0                   0
## 6          AND        Andorra                     0                   0
##   emissions.extraction emissions.manufacturing emissions.maritime
## 1                 1130                 1822167                  0
## 2                    0                       0                  0
## 3                11434                 2465018              30560
## 4                97162                25841289             455958
```

```
## 5                    0                     353                 0
## 6                    0                       0                 0
##   emissions.oil and gas emissions.power emissions.transport emissions.waste
## 1                522846         3742644              2433212         5824000
## 2                     0               0                    0               0
## 3                 65166               0              2039538          470010
## 4              81920699        41634720             46011348        19434645
## 5                     0          138000                 5195           10604
## 6                     0           21000                    0               0
##   emissions.total
## 1        30467148
## 2               0
## 3         8704608
## 4       254223934
## 5          154152
## 6           21000
```

```
filename_gdp = paste(data_path, "GDP_worldbank.xls", sep = "")
# Reading data, Renaming variables to the correct names in the third row, Dropping the first three irre
gdp = read_excel(filename_gdp, skip = 3, col_names = TRUE)

# Rename the 2020 GDP variable simply to GDP and Country code to match emission data
names(gdp)[names(gdp) == "2020"] <- "GDP"
names(gdp)[names(gdp) == "Country Code"] <- "country code"
# Drop all irrelevant variables
gdp = gdp[c("country code", "GDP")]
gdp$GDP = as.numeric(gdp$GDP)

df_per_sector = merge(per_sector_emissions, gdp, by = "country code", all = F)

# Saving dataframe as csv file
filename_per_sector_gdp = paste(data_path, "/output/per_sector_emissions_gdp.csv", sep = "")
write.csv(df_per_sector, file = filename_per_sector_gdp)
```

2. Adding 2020 GDP data from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2021&start=1960

```
filename_pledges = paste(data_path, "/pledges/net-zero-targets.csv", sep = "")
pledges = read.csv(filename_pledges,sep = ",", header = TRUE)
names(pledges)[names(pledges) == "Code"] <- "country code"
names(pledges)[names(pledges) == "Year"] <- "net_zero_target"
pledges = pledges[c("country code", "net_zero_target")]

df_pledges = merge(df_per_sector, pledges, by = "country code", all = F)
rownames(df_pledges) = df_pledges[,1]
head(df_pledges)
```

3. Merging data on Climate Pledges from https://ourworldindata.org/grapher/net-zero-targets?country=SOM~BRA~MDG

```
##     country code         country name emissions.agriculture emissions.buildings
## ARE          ARE United Arab Emirates               1952655              932653
## ATG          ATG  Antigua and Barbuda                 19536               88431
```

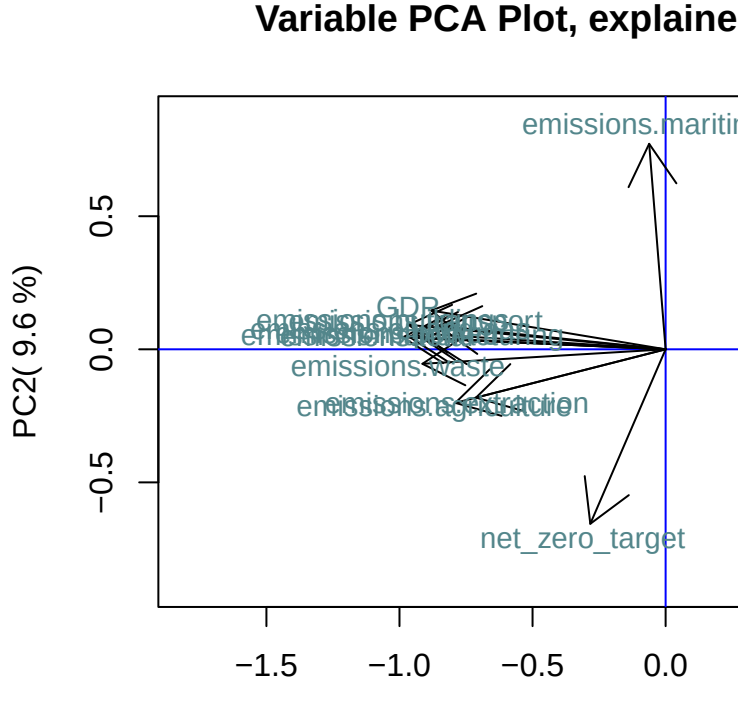
```
## AUS          AUS            Australia              93785962            43380686
## AUT          AUT              Austria               7093716            10952267
## BEN          BEN                Benin               5726903              711837
## BHR          BHR              Bahrain                 72994              275711
##     emissions.extraction emissions.manufacturing emissions.maritime
## ARE               119524                82332009            134440
## ATG                   51                   38824          10338235
## AUS              8292742                45058599            752974
## AUT                65771                17576600                 0
## BEN                 1470                 1170073                 0
## BHR                    0                 5478661             57497
##     emissions.oil and gas emissions.power emissions.transport emissions.waste
## ARE              74389570        84635680            47904341         7495600
## ATG                     0          275000              288564           49280
## AUS             129382650       180107320           115295903        13944420
## AUT               2566065        12342280            21271638         2814753
## BEN               1327200          153233             6830284         2464354
## BHR              10211411        20961760             4317051          996800
##     emissions.total          GDP net_zero_target
## ARE       299896472 3.588688e+11            2050
## ATG        11097921 1.370281e+09            2040
## AUS       630001256 1.327836e+12            2050
## AUT        74683090 4.332585e+11            2040
## BEN        18385354 1.565155e+10            2000
## BHR        42371885 3.472336e+10            2060
```

```r
# Saving dataframe as csv file
filename_merged_pledges= paste(data_path, "/output/pledges_emissions_gdp.csv", sep = "")
write.csv(df_pledges, file = filename_merged_pledges)
```

```r
# Transforming dataframe into matrix to perform matrix operations
X = as.matrix(df_pledges)
# Getting rid of non numercal variables
rownames(X) = X[,1]
X = X[,c(-1,-2)]
X = apply(X, 2, as.numeric)

# Scaling data
Xs = scale(X)
# Creating correlation matrix R
R = cor(Xs)
# Eigenvectors and values
E = eigen(R)$vectors
l = eigen(R)$values

R2 = l / sum(l) * 100
# The first two principal components explain 77% of the variance

# Perform the PCA
Y = Xs %*% E
colnames(Y) <- paste0("PC", 1:ncol(Y))
W = cor(X,Y)
```

4. Performing the Principle Component Analyis manually for practice

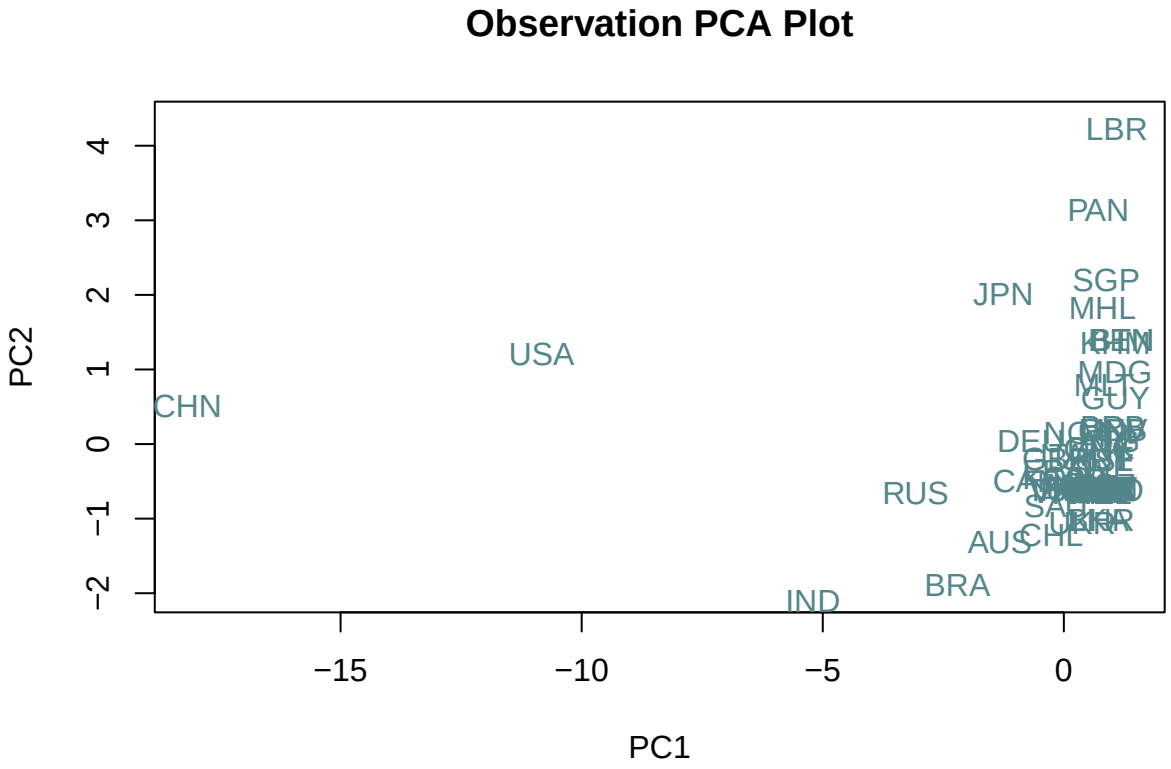
```
# Limits for the graph with the variables
inferior = min(W[,1:2])
superior = max(W[,1:2])
limits = 1.01 * c(inferior, superior) + 0.1

# Defining the limits and the titles for the graph
plot(W[,1:2],
  type="n",
  asp = 1,
  main=paste("Variable PCA Plot, explained Variance is ",
  round(R2[1]+R2[2],1), "%"),
  xlim = limits,
  ylim = limits,
  xlab=paste("PC1(", round(R2[1],1), "%)"),
  ylab=paste("PC2(", round(R2[2],1), "%)")
)

# The axis
abline(v=0, col="blue")
abline(h=0, col="blue")

# Drawing the arrows
arrows(0,0,W[,1],W[,2])

# Adding the labels
text(1.1 * W[,1:2], colnames(X), col= "cadetblue4",cex=0.9)
```



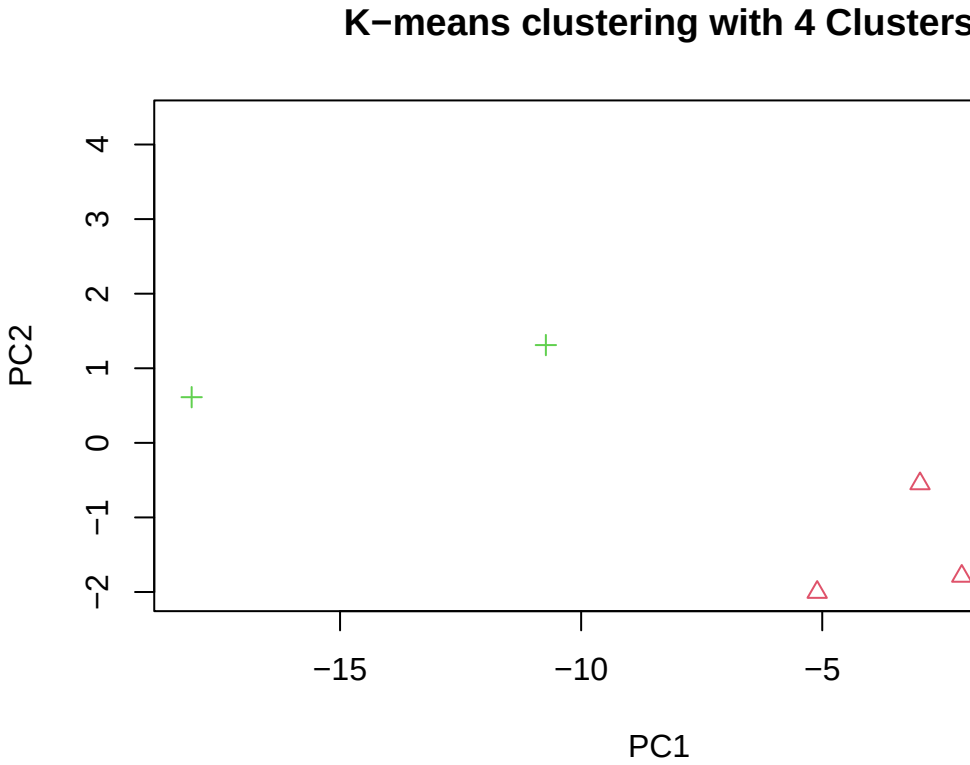
5. Plotting Variables and Observations in a PCA Plot

- The first principal component is largely determined by GDP and general emissions (without maritime emissions)
- The second principal component is mostly determined by the net-zero-target and by maritime emissions

```
# Plotting observations
plot(Y[,1],Y[,2],xlab='PC1',ylab='PC2', type = "n", main=paste("Observation PCA Plot"))
points(Y[,1:2], pch="")
text(Y[,1]-0.1,Y[,2]-0.1, labels=rownames(df_pledges), col = "cadetblue4")
```




```
# Chose k = 4 clusters
km <- kmeans(Xs, centers = 4)
clus <- km$cluster
# Plotting Clusters
plot(Y[,1],Y[,2],col=clus,pch=clus,xlab='PC1',ylab='PC2',main="K-means clustering with 4 Clusters")
text(Y[,1]-0.1,Y[,2]-0.1, labels=rownames(X))
```



6. Performing K-Means Clustering

- China and the US form a cluster in most iterations and Indonesia, Brazil, Russia and Australia form another cluster
- To gain a better understanding of emission profiles, more data must be added