WVS

Marek Chadim

2024-09-24

1 Download data and documentation

The dataset and the codebook from the World Value Survey can be downloaded from http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp (http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp)

The data has $89{,}565$ rows and 442 columns.

```
dim(WV6_Data_R_v20201117)
```

```
## [1] 89565   442
```

2 Creating the data set

```r
# Create the country variable based on country codes
d <- WV6_Data_R_v20201117
d <- d |> mutate(country = countrycode(V2, origin = "iso3n", destination = "country.name"))

# View the country variable to verify correct labeling
View(d$country)

# Select variables related to values, social trust, happiness, and political orientation
selected_vars <- c("V5", "V6", "V7", "V8", "V9", "V10", "V11", "V23", "V24",
                   "V55", "V56", "V59", "V67", "V69", "V70", "V71", "V72",
                   "V73", "V74", "V76", "V77", "V78", "V79", "V102", "V103",
                   "V104", "V105", "V106", "V107", "V108", "V109", "V110",
                   "V111", "V112", "V113", "V114", "V115", "V116", "V117",
                   "V118", "V119", "V120", "V121", "V122", "V123", "V124",
                   "V127", "V128", "V130", "V131", "V132", "V133", "V134",
                   "V135", "V136", "V137", "V138", "V139", "V97", "V98",
                   "V99", "V100", "V101", "V140", "V141", "V142", "V143",
                   "V145", "V146", "V147", "V148", "V149", "V150", "V151",
                   "V152", "V153", "V154", "V155", "V156", "V157", "V158",
                   "V159", "V160", "V161", "V162", "V163", "V164", "V165",
                   "V166", "V167", "V168", "V169", "V170")

# Subset the data to the selected variables
d <- d |> select(country, all_of(selected_vars))

# Handle missing values by replacing specific negative codes with NA
d <- d |>
  mutate(across(where(is.numeric), ~na_if(., -1))) |>
  mutate(across(where(is.numeric), ~na_if(., -2))) |>
  mutate(across(where(is.numeric), ~na_if(., -3))) |>
  mutate(across(where(is.numeric), ~na_if(., -4))) |>
  mutate(across(where(is.numeric), ~na_if(., -5)))

# Aggregate data to the country level by calculating the mean of all variables
d <- d |> group_by(country) |> summarise(across(where(is.numeric), ~mean(.x, na.rm = TRUE)))

# Remove any columns with missing data
d <- d[, colSums(is.na(d)) == 0]
colSums(is.na(d)) == 0  # Check to confirm no missing values remain
```

```
## country      V5      V6      V7      V8      V9     V10     V11     V23     V24
##    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE
##     V55     V59     V67     V69     V70     V71     V72     V73     V76     V77
##    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE
##     V78     V79    V108    V110    V111    V113    V114    V115    V117    V119
##    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE
##    V120    V121    V122    V123    V124    V131    V132    V133    V134    V136
##    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE
##    V137    V138    V139     V97     V98     V99    V100    V101    V140    V143
##    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE
##    V150    V151    V153    V155    V156    V170
##    TRUE    TRUE    TRUE    TRUE    TRUE    TRUE
```

```
# Write the cleaned data to a CSV file
write.csv(d, "wvs.csv", row.names = FALSE)
```

3 Analysis

PCA

PCA is performed on the dataset after scaling the variables to have a standard deviation of one.

```
wvs <- read.csv("wvs.csv")
row.names(wvs) <- wvs[,1]
wvs <- wvs[,-1]
countries <- row.names(wvs)
countries
```

```
##  [1] "Algeria"                "Argentina"
##  [3] "Armenia"                "Australia"
##  [5] "Azerbaijan"             "Belarus"
##  [7] "Brazil"                 "Chile"
##  [9] "China"                  "Colombia"
## [11] "Cyprus"                 "Ecuador"
## [13] "Egypt"                  "Estonia"
## [15] "Georgia"                "Germany"
## [17] "Ghana"                  "Haiti"
## [19] "Hong Kong SAR China"    "India"
## [21] "Iraq"                   "Japan"
## [23] "Jordan"                 "Kazakhstan"
## [25] "Kuwait"                 "Kyrgyzstan"
## [27] "Lebanon"                "Libya"
## [29] "Malaysia"               "Mexico"
## [31] "Morocco"                "Netherlands"
## [33] "New Zealand"            "Nigeria"
## [35] "Pakistan"               "Palestinian Territories"
## [37] "Peru"                   "Philippines"
## [39] "Poland"                 "Qatar"
## [41] "Romania"                "Russia"
## [43] "Rwanda"                 "Singapore"
## [45] "Slovenia"               "South Africa"
## [47] "South Korea"            "Spain"
## [49] "Sweden"                 "Taiwan"
## [51] "Thailand"               "Trinidad & Tobago"
## [53] "Tunisia"                "Turkey"
## [55] "Ukraine"                "United States"
## [57] "Uruguay"                "Uzbekistan"
## [59] "Yemen"                  "Zimbabwe"
```

```
names(wvs)
```

```
##  [1] "V5"   "V6"   "V7"   "V8"   "V9"   "V10"  "V11"  "V23"  "V24"  "V55" 
## [11] "V59"  "V67"  "V69"  "V70"  "V71"  "V72"  "V73"  "V76"  "V77"  "V78" 
## [21] "V79"  "V108" "V110" "V111" "V113" "V114" "V115" "V117" "V119" "V120"
## [31] "V121" "V122" "V123" "V124" "V131" "V132" "V133" "V134" "V136" "V137"
## [41] "V138" "V139" "V97"  "V98"  "V99"  "V100" "V101" "V140" "V143" "V150"
## [51] "V151" "V153" "V155" "V156" "V170"
```

Without normalization, results differ due to variable scale differences:

```
apply(wvs, 2, mean)
```

```
##       V5       V6       V7       V8       V9      V10      V11      V23 
## 1.673468 1.887062 2.643526 1.508577 1.871907 1.861636 2.096946 6.844663 
##      V24      V55      V59      V67      V69      V70      V71      V72 
## 1.761284 7.115627 5.883512 2.306926 1.483234 2.764252 3.809883 2.340996 
##      V73      V76      V77      V78      V79     V108     V110     V111 
## 3.176624 3.740317 2.517224 2.501468 2.486355 2.157033 2.648706 2.539744 
##     V113     V114     V115     V117     V119     V120     V121     V122 
## 2.404164 2.444360 2.595923 2.760835 2.153863 2.528054 2.442902 2.398329 
##     V123     V124     V131     V132     V133     V134     V136     V137 
## 2.414271 2.331160 6.276717 4.214201 8.066850 7.014702 7.418509 5.950238 
##     V138     V139      V97      V98      V99     V100     V101     V140 
## 6.025539 7.892392 5.630597 4.474456 3.808070 4.180620 6.317422 8.318191 
##     V143     V150     V151     V153     V155     V156     V170 
## 1.813585 1.694779 1.691183 2.393354 2.613034 2.168286 1.887675
```

```
apply(wvs, 2, var)
```

```
##         V5         V6         V7         V8         V9        V10        V11 
## 0.03703609 0.06066860 0.08015831 0.05343684 0.48249065 0.07118806 0.07739582 
##        V23        V24        V55        V59        V67        V69        V70 
## 0.61721848 0.02639433 0.43928619 0.76714587 0.09106552 0.10271962 0.18340981 
##        V71        V72        V73        V76        V77        V78        V79 
## 0.44105398 0.19582324 0.38651597 0.23996851 0.17834372 0.17537272 0.33629480 
##       V108       V110       V111       V113       V114       V115       V117 
## 0.26287357 0.09323242 0.09637715 0.15511457 0.16313199 0.16737395 0.18586148 
##       V119       V120       V121       V122       V123       V124       V131 
## 0.05984647 0.06609448 0.12002990 0.06999929 0.09671995 0.07437698 0.85149690 
##       V132       V133       V134       V136       V137       V138       V139 
## 1.58755487 0.48296989 0.78044578 0.42980964 1.39805265 1.60668291 0.64561039 
##        V97        V98        V99       V100       V101       V140       V143 
## 0.51037516 1.02640530 0.55346156 0.95605035 0.82087795 0.25029453 0.06612051 
##       V150       V151       V153       V155       V156       V170 
## 0.04119235 0.05638046 0.47716279 0.18535672 0.08162220 0.10786429
```

Running PCA with scaling:

```
pr.out <- prcomp(wvs, scale = TRUE)
names(pr.out)
```

```
## [1] "sdev"     "rotation" "center"   "scale"    "x"
```

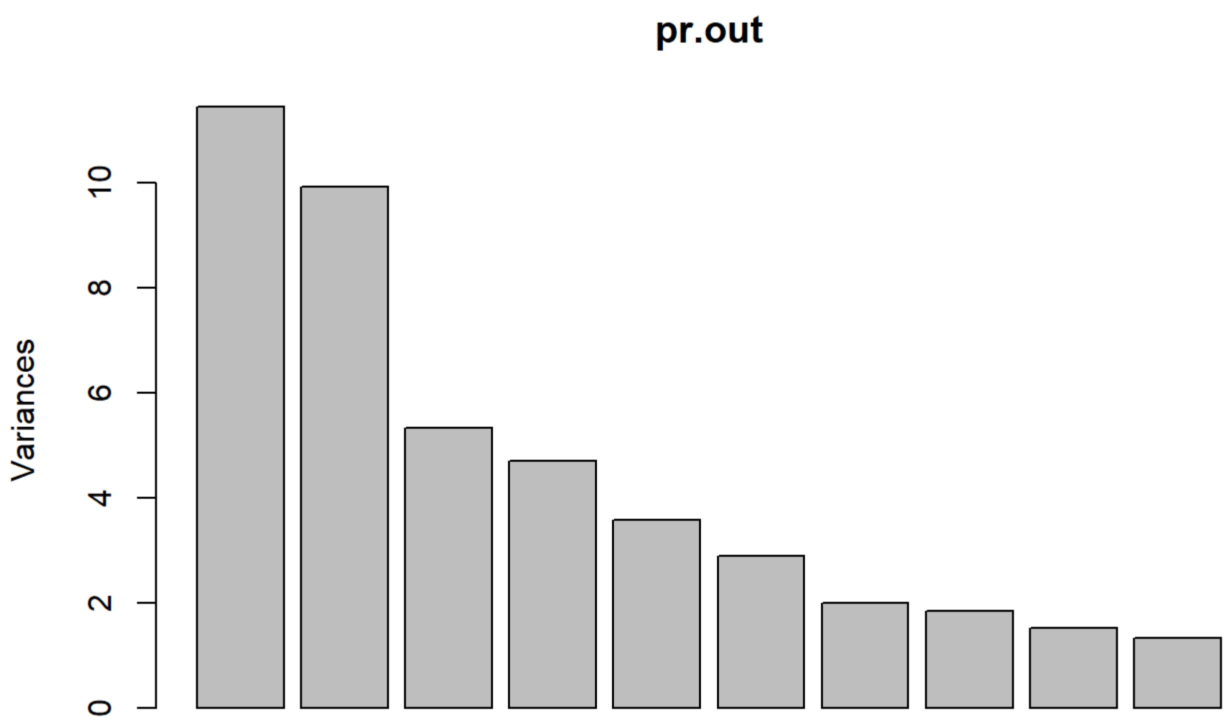
```
dim(pr.out$x)
```

```
## [1] 60 55
```

```
summary(pr.out)
```

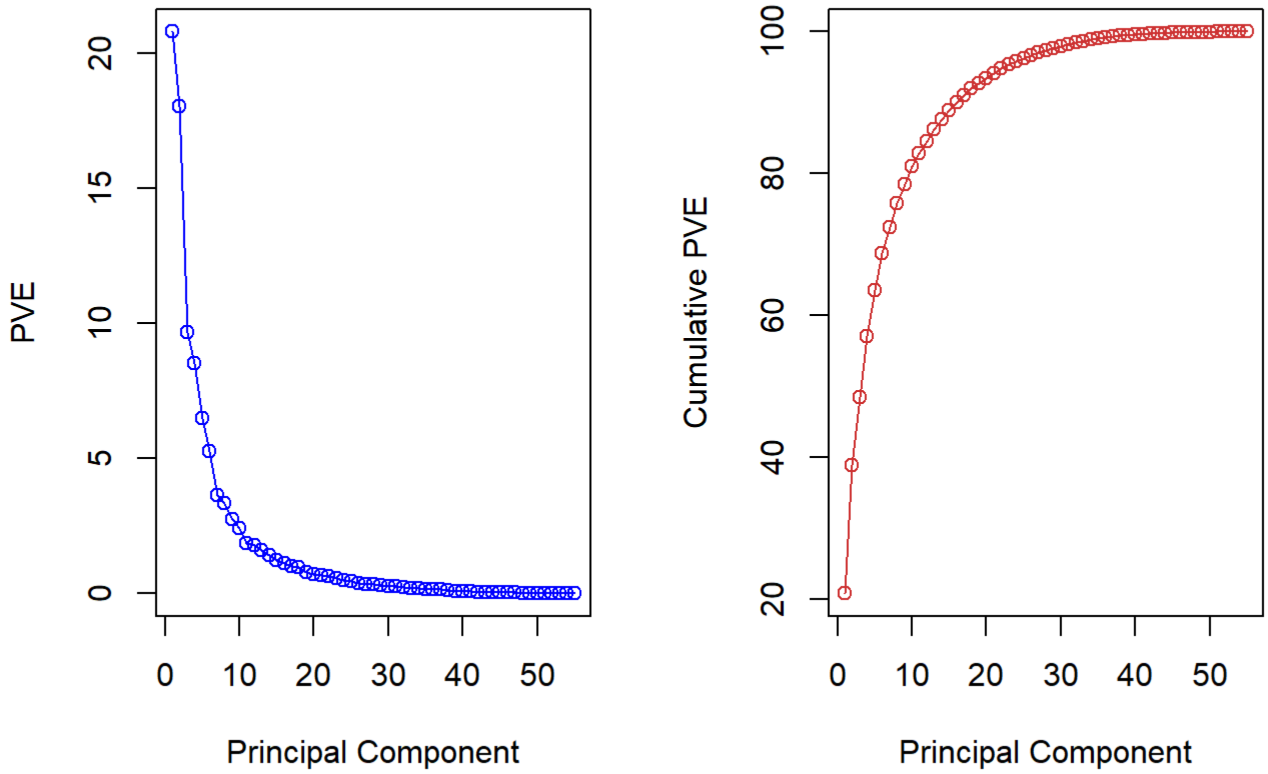
```
## Importance of components:
##                           PC1    PC2     PC3     PC4     PC5     PC6     PC7
## Standard deviation     3.3838 3.1496 2.30784 2.16667 1.88909 1.70280 1.41208
## Proportion of Variance 0.2082 0.1804 0.09684 0.08535 0.06488 0.05272 0.03625
## Cumulative Proportion  0.2082 0.3886 0.48539 0.57074 0.63563 0.68835 0.72460
##                            PC8     PC9    PC10    PC11    PC12    PC13    PC14
## Standard deviation     1.35511 1.23054 1.15231 1.00709 0.99471 0.94145 0.88534
## Proportion of Variance 0.03339 0.02753 0.02414 0.01844 0.01799 0.01612 0.01425
## Cumulative Proportion  0.75799 0.78552 0.80966 0.82810 0.84609 0.86221 0.87646
##                          PC15    PC16    PC17    PC18    PC19    PC20    PC21
## Standard deviation     0.8225 0.79028 0.75042 0.72626 0.65343 0.63210 0.61054
## Proportion of Variance 0.0123 0.01136 0.01024 0.00959 0.00776 0.00726 0.00678
## Cumulative Proportion  0.8888 0.90011 0.91035 0.91994 0.92771 0.93497 0.94175
##                           PC22    PC23    PC24    PC25    PC26    PC27    PC28
## Standard deviation     0.59652 0.55076 0.52114 0.49061 0.45879 0.43423 0.43168
## Proportion of Variance 0.00647 0.00552 0.00494 0.00438 0.00383 0.00343 0.00339
## Cumulative Proportion  0.94822 0.95373 0.95867 0.96305 0.96687 0.97030 0.97369
##                           PC29    PC30    PC31    PC32    PC33    PC34   PC35
## Standard deviation     0.40129 0.39337 0.38207 0.36615 0.33833 0.33273 0.3060
## Proportion of Variance 0.00293 0.00281 0.00265 0.00244 0.00208 0.00201 0.0017
## Cumulative Proportion  0.97662 0.97943 0.98209 0.98452 0.98660 0.98862 0.9903
##                           PC36    PC37    PC38    PC39    PC40    PC41    PC42
## Standard deviation     0.28798 0.28207 0.23865 0.22115 0.21409 0.20847 0.17683
## Proportion of Variance 0.00151 0.00145 0.00104 0.00089 0.00083 0.00079 0.00057
## Cumulative Proportion  0.99183 0.99327 0.99431 0.99520 0.99603 0.99682 0.99739
##                           PC43    PC44    PC45    PC46    PC47   PC48    PC49
## Standard deviation     0.16221 0.15270 0.14634 0.14085 0.12319 0.1058 0.09576
## Proportion of Variance 0.00048 0.00042 0.00039 0.00036 0.00028 0.0002 0.00017
## Cumulative Proportion  0.99787 0.99829 0.99868 0.99904 0.99932 0.9995 0.99969
##                           PC50    PC51    PC52    PC53    PC54    PC55
## Standard deviation     0.08156 0.06723 0.05291 0.04476 0.02539 0.02074
## Proportion of Variance 0.00012 0.00008 0.00005 0.00004 0.00001 0.00001
## Cumulative Proportion  0.99981 0.99989 0.99994 0.99998 0.99999 1.00000
```

```
plot(pr.out)
```



A scree plot shows the PVE by each principal component, with the first four explaining over 50% of the variance:

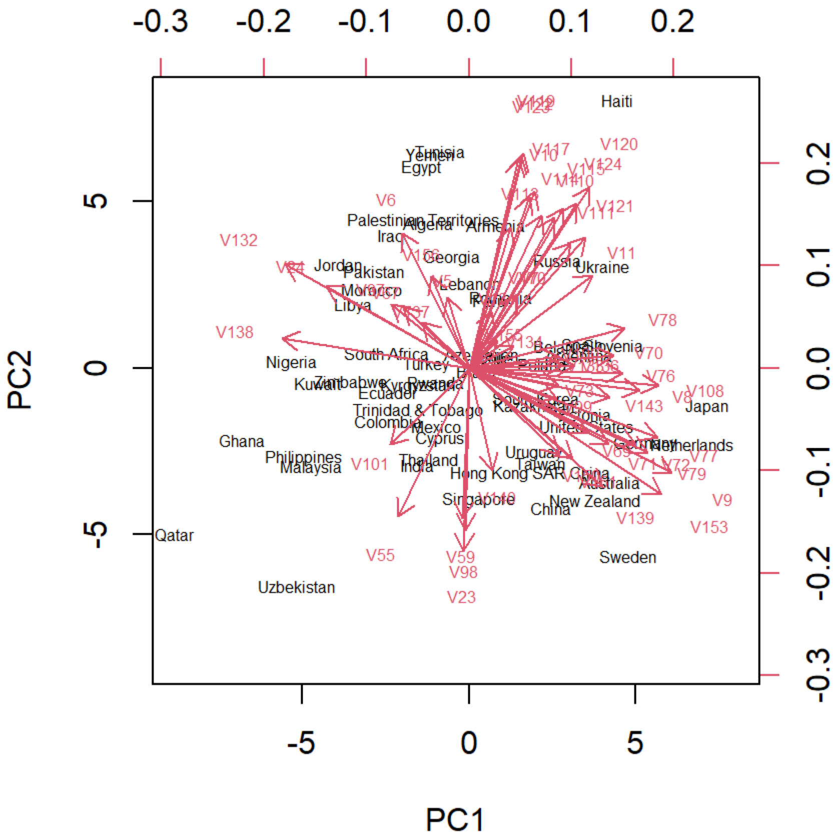
```
pve <- 100 * pr.out$sdev^2 / sum(pr.out$sdev^2)
par(mfrow = c(1, 2))
plot(pve,  type = "o", ylab = "PVE",
    xlab = "Principal Component", col = "blue")
plot(cumsum(pve), type = "o", ylab = "Cumulative PVE",
    xlab = "Principal Component", col = "brown3")
```



Cultural Map

A biplot visualizes countries in the principal component space:

```
par(mfrow = c(1, 1))
biplot(pr.out, scale = 0, cex=.5)
```



Interpretation

PC1 (Self-expression vs. Survival values): This axis likely reflects the degree to which countries emphasize individual freedoms, self-expression, and creativity (variables such as V70 “New ideas and creativity”). Countries scoring high on this axis, like Sweden and Germany, represent liberal democracies with a high emphasis on self-expression and secular values. In contrast, countries scoring lower, such as Uzbekistan and Qatar, tend to prioritize survival values such as tradition and authority .

PC2 (Economic vs. Traditional values): The second component seems to be linked to economic ideology and traditional beliefs. Variables such as V132 (“Religious authorities interpret laws”) and V97 (“Private vs. state ownership”) load heavily on this component, indicating a spectrum from market-driven economies to more traditionally governed societies. Japan and Sweden score high on this component, reflecting their advanced economies and progressive values, while countries like Haiti and Egypt cluster together, possibly due to their emphasis on traditional authority structures

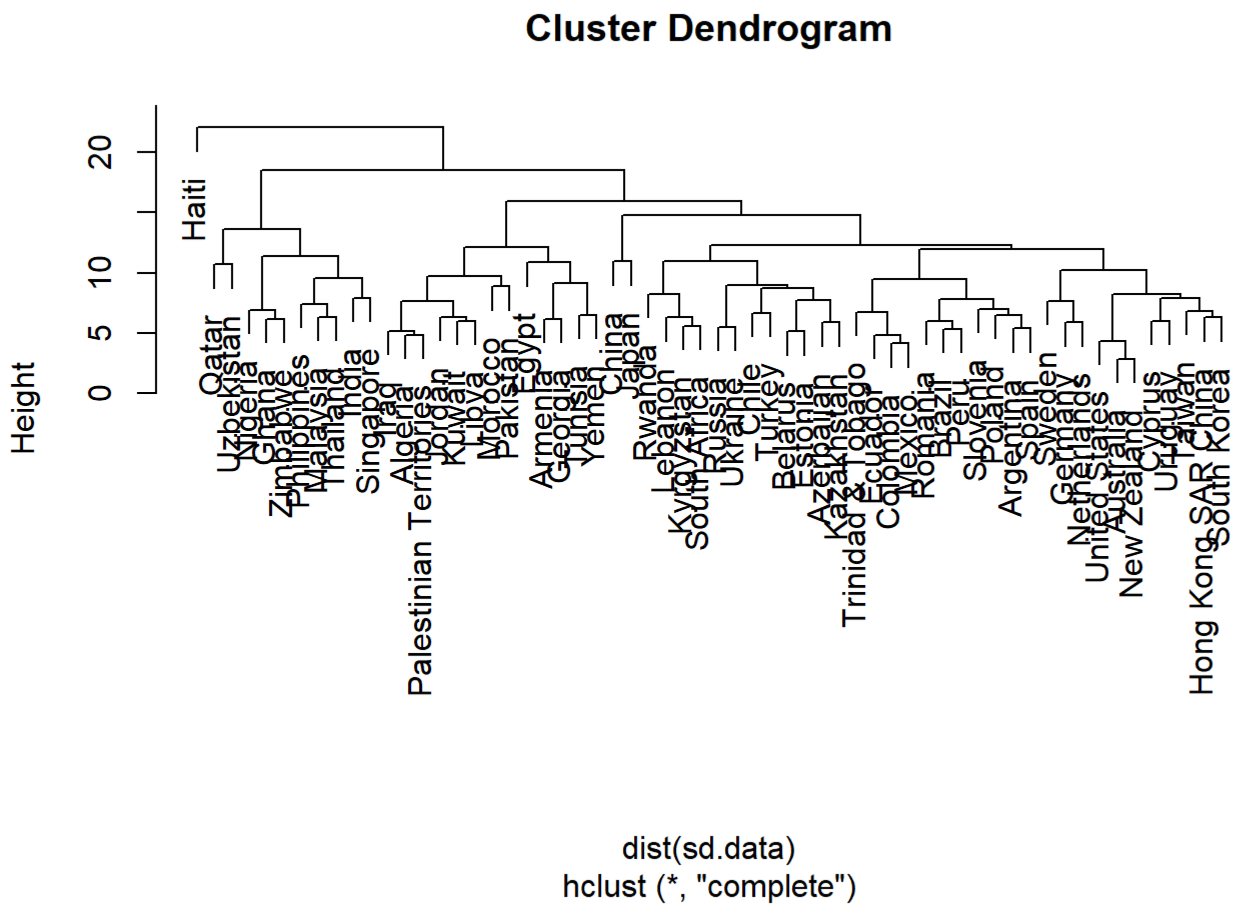
The analysis confirms the importance of self-expression versus survival values along the first principal component (PC1). However, the second principal component (PC2) does not clearly align with the secular-rational versus traditional divide. Instead, it seems to capture a spectrum of social hierarchy, religious authority, and respect for governance. Countries like Haiti and Egypt score high on PC2, reflecting the role of religious influence and social obedience, whereas countries like Japan and Sweden score lower, signaling a preference for secular, rational governance with less hierarchical structures.

Clustering Analysis of Countries

```
wvs <- wvs[,-1]
sd.data <- scale(wvs)
hc.out <- hclust(dist(sd.data))
hc.clusters <- cutree(hc.out, 4)
table(hc.clusters, row.names(wvs))
```

```
## 
## hc.clusters Algeria Argentina Armenia Australia Azerbaijan Belarus Brazil Chile
##           1       1         0       1         0          0       0      0     0
##           2       0         1       0         1          1       1      1     1
##           3       0         0       0         0          0       0      0     0
##           4       0         0       0         0          0       0      0     0
## 
## hc.clusters China Colombia Cyprus Ecuador Egypt Estonia Georgia Germany Ghana
##           1     0        0      0       0     1       0       1       0     0
##           2     1        1      1       1     0       1       0       1     0
##           3     0        0      0       0     0       0       0       0     1
##           4     0        0      0       0     0       0       0       0     0
## 
## hc.clusters Haiti Hong Kong SAR China India Iraq Japan Jordan Kazakhstan Kuwait
##           1     0                   0     0    1     0      1          0      1
##           2     0                   1     0    0     1      0          1      0
##           3     0                   0     1    0     0      0          0      0
##           4     1                   0     0    0     0      0          0      0
## 
## hc.clusters Kyrgyzstan Lebanon Libya Malaysia Mexico Morocco Netherlands
##           1          0       0     1        0      0       1           0
##           2          1       1     0        0      1       0           1
##           3          0       0     0        1      0       0           0
##           4          0       0     0        0      0       0           0
## 
## hc.clusters New Zealand Nigeria Pakistan Palestinian Territories Peru
##           1           0       0        1                       1    0
##           2           1       0        0                       0    1
##           3           0       1        0                       0    0
##           4           0       0        0                       0    0
## 
## hc.clusters Philippines Poland Qatar Romania Russia Rwanda Singapore Slovenia
##           1           0      0     0       0      0      0         0        0
##           2           0      1     0       1      1      1         0        1
##           3           1      0     1       0      0      0         1        0
##           4           0      0     0       0      0      0         0        0
## 
## hc.clusters South Africa South Korea Spain Sweden Taiwan Thailand
##           1            0           0     0      0      0        0
##           2            1           1     1      1      1        0
##           3            0           0     0      0      0        1
##           4            0           0     0      0      0        0
## 
## hc.clusters Trinidad & Tobago Tunisia Turkey Ukraine United States Uruguay
##           1                 0       1      0       0             0       0
##           2                 1       0      1       1             1       1
##           3                 0       0      0       0             0       0
##           4                 0       0      0       0             0       0
## 
## hc.clusters Uzbekistan Yemen Zimbabwe
##           1          0     1        0
##           2          0     0        0
##           3          1     0        1
##           4          0     0        0
```

```
par(mfrow = c(1, 1))
plot(hc.out, labels = row.names(wvs))
abline(h = 139, col = "red")
```



Interpretation of Country Clusters:

Distinct Cluster - Haiti: Haiti stands out at a high level in the hierarchy, suggesting its responses are notably distinct, likely due to unique social, political, or cultural traits.

Western Countries: The USA, Germany, Sweden, New Zealand, Australia, and the Netherlands form a close cluster, reflecting high-income, industrialized nations with shared values of individualism, democracy, secularism, and human rights.

Latin American and European Countries: Argentina, Brazil, Chile, Peru, Spain, and Slovenia cluster together, likely blending European heritage with Latin American values focused on community, family, and moderate economic priorities.

Middle Eastern and African Countries: Pakistan, Kuwait, Iraq, Libya, South Africa, Nigeria, and Malaysia form a broader cluster, reflecting traditionalism, strong family values, and religious orientations.

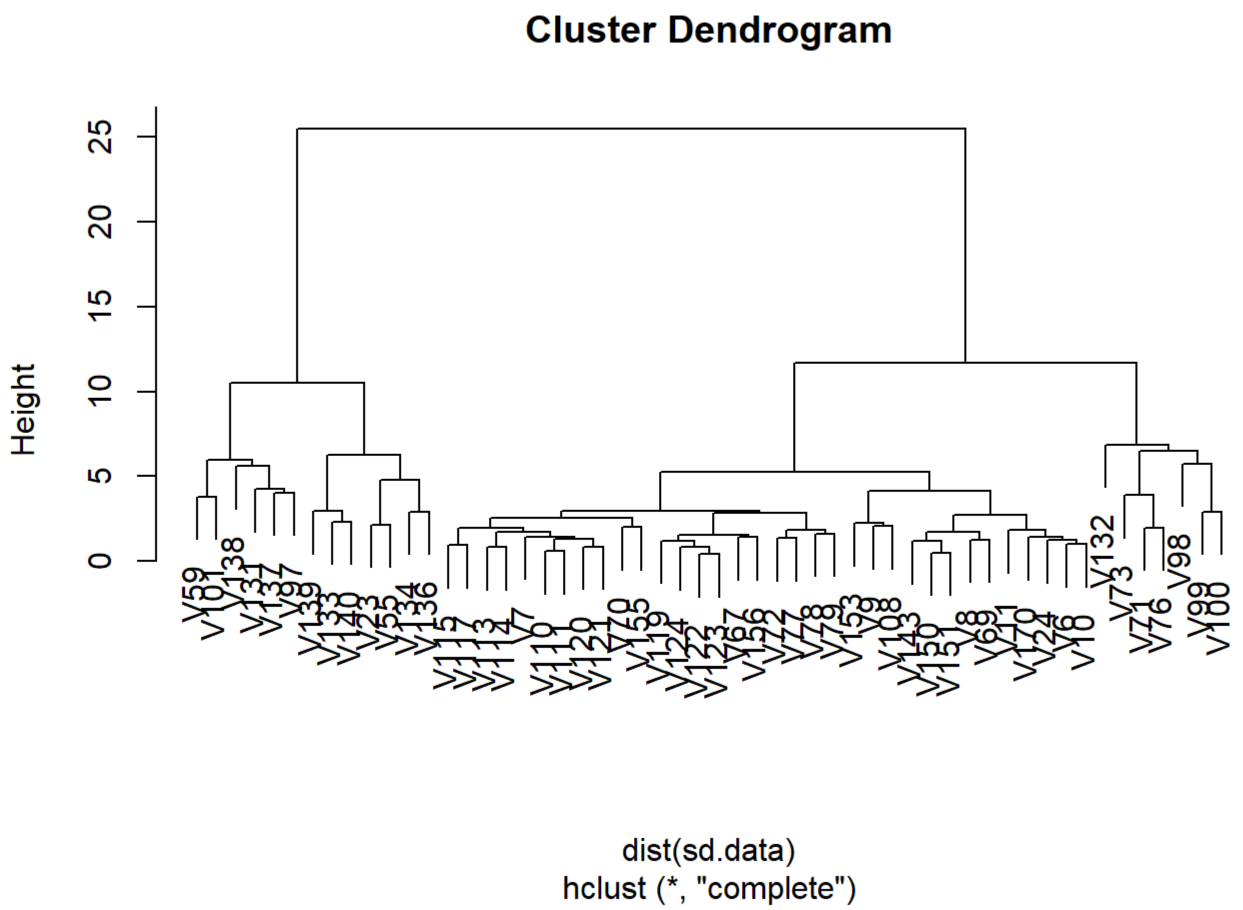
Central/Eastern Europe and Asia: Russia, Kazakhstan, Azerbaijan, and China cluster, possibly due to post-Soviet or authoritarian influences, sharing similar values around authority and governance.

Clustering Analysis of Variables

```
wvs.data <- t(wvs)
sd.data <- scale(wvs.data)
hc.out <- hclust(dist(sd.data))
hc.clusters <- cutree(hc.out, 4)
table(hc.clusters, colnames(wvs))
```

```
## 
## hc.clusters V10 V100 V101 V108 V11 V110 V111 V113 V114 V115 V117 V119 V120 V121
##           1   1    0    0    1   1    1    1    1    1    1    1    1    1    1
##           2   0    0    0    0   0    0    0    0    0    0    0    0    0    0
##           3   0    0    1    0   0    0    0    0    0    0    0    0    0    0
##           4   0    1    0    0   0    0    0    0    0    0    0    0    0    0
## 
## hc.clusters V122 V123 V124 V131 V132 V133 V134 V136 V137 V138 V139 V140 V143
##           1    1    1    1    0    0    0    0    0    0    0    0    0    1
##           2    0    0    0    0    0    1    1    1    0    0    1    1    0
##           3    0    0    0    1    0    0    0    0    1    1    0    0    0
##           4    0    0    0    0    1    0    0    0    0    0    0    0    0
## 
## hc.clusters V150 V151 V153 V155 V156 V170 V23 V24 V55 V59 V6 V67 V69 V7 V70 V71
##           1    1    1    1    1    1    1   0   1   0   0  1   1   1  1   1   0
##           2    0    0    0    0    0    0   1   0   1   0  0   0   0  0   0   0
##           3    0    0    0    0    0    0   0   0   0   1  0   0   0  0   0   0
##           4    0    0    0    0    0    0   0   0   0   0  0   0   0  0   0   1
## 
## hc.clusters V72 V73 V76 V77 V78 V79 V8 V9 V97 V98 V99
##           1   1   0   0   1   1   1  1  1   0   0   0
##           2   0   0   0   0   0   0  0  0   0   0   0
##           3   0   0   0   0   0   0  0  0   1   0   0
##           4   0   1   1   0   0   0  0  0   0   1   1
```

```
par(mfrow = c(1, 1))
plot(hc.out, labels = colnames(wvs))
abline(h = 139, col = "red")
```



Interpretation of Variable Clusters:

Cluster 1: This cluster contains variables related to financial satisfaction and democratic ideals, indicating an economic-political ideology theme.

Cluster 2: Comprises variables focused on personal values, security, and creativity, balancing traditional and innovative aspects.

Cluster 3: Primarily includes economic ideology questions such as private vs. state ownership and competition, highlighting ideological differences.

Cluster 4: Features variables emphasizing democratic values and governance structures, particularly regarding the role of religion and elections.