DSCI300 Mini Project 3 - Tidy Data and Code Books

Nisi Mohan Kuniyil 300321388

2020-10-21

1 CODEBOOK

1.1 Rawdata (dem_score.csv)

1. The dataset contains observations of democracy index of countries from 1952 to 1992.

2. There are 96 rows and 10 columns.

3. The first column records country names, and second onward, columns for years starting from 1952 to 1992, in an interval of 5 each showing democracy index of countries for that particular year.

4. Each row captures the democracy index of a specific country over the years.

5. Class of the first column country is "char", rest of the columnse are "int".

6. The democracy index value is in a scale from -10 to 10, 10 representing highly democratic and -10 represents highly autocratic nations.

7. As per Wikipedia's "Democacy Index", the distribution of index starts from 0 to 10 and it has been categorised into four levels.
   - Countries which fall into [0-4] bin is considered to be under authoritarian regimes.
   - [4.01-6] under Hybrid regimes.
   - [6.01-8] under flawed democracies.
   - [8.01-10] under full democracies.
8. Using this information, the index values in the raw data can also be split into four levels.

- [-10, -5) falls under authoritarian regimes.
- [-5,0) into hybrid regimes.
- [0,5) can be called as flawed democracies, and
- [5,10] as full democracies.

1.2 Cleaned dataset (demscoreNew)

1. The raw dataset is converted to Long format.

2. The new format contains three columns (Country, Year, DemIndex), and a total of 864 observations.

3. Column Country is a class of "chr", Year is class of "fctr", and DemIndex is "int".

4. DemIndex contains the democracy index values ranging from -10 to 10.

5. Country column contains country names and the Year column contains years over the period 1952 to 1992.

1.3 Other datasets used for better data visualization

1.3.1 world dataset from ggplot2.map_data for plotting map visualization

1. World_map is taken from library ggplot2 using the function map_data.

2. There are 99338 observations and 6 columns, which are long, lat, group, order, region, and subregion in the world_map dataset.

3. Classes of long, lat, and group are "num", order is "int", region and subregion are "chr".

1.3.2 country codes dataset from raster package

1. ccodes() from raster package is usedfor getting ISO2 code for corresponding countries in the given dataset.

2. There are 256 observations and 10 columns, which are Name, ISO3, ISO2, Name_ISO, Name_FAO, Name_Local, Sovereign, Unregion1, unregion2, and continent in the ccodes() dataframe.

3. Classes of all the above columns are "chr".

1.4 Tidying the dem_score.csv data frame and performing other operations for better visualization:

1. The row dataframe is in wide format, in order to analyze the data in an effective way it has to be changed from wide to long format.

2. To perform this operation, use "melt" function, resulting dataframe is "moltendf".

3. Then use "moltendf" and rename the column "value" to DemIndex and assign that to a new data set "demscoreNew".

4. Check for NA's in the new dataframe "demscoreNew".

5. According to Wiki, democracy index has four levels, based on this "DemIndex" is also split into four levels for easier analysis and visualization.

6. These levels are added as a new column "Interval" in "demscoreNew". Interval is a class of factor.

7. Taking the count of countries in each level by grouping by "Year" and "Interval" to plot a stacked bar plot to show the democracy index level distribution over all countries during 1952 to 1992 and assigning this to a new variable " demCount".

8. Using the "demCount" data frame, use ggplot to create a stacked bar plot.

9. Because of the name mismatch of countries between world_map, ccodes(), and dem_score data frames, a new column"ISO2" is added converting country names into standard ISO2 code.

10. Merging demscoreNew with ccodes() to get continent info.

11. Creating function "plot_dem_on_map" to plot democracy index levels on a map.
    - The function takes arguments df: which is data frame and year.
      - function signature: plot_dem_on_map(df, year).

    - Function steps
      - Filter the dataframe based on the year passed, assigning that to a variable "demMap".

      - Get world map data using map_data function in ggplot2 and assigning it to world_map.

      - Add ISO2 column to the world map data for merging with demMap.

      - Left join demMap with world_map by column ISO2.

      - Use ggplot and geom_polygon to plot democracy index levels on map.
12. Now the function can be called multiple times for any year we want to plot the color coded map.

13. Merge demScoreNew dataframe with ccode() to get the continent information.

14. Taking average of DemIndex to plot average democracy index of each continent over the period 1952-1992 and assigning this to a variable: " demIndexAverage".

15. Using the demIndexAverage data frame, plotting multiple line plots for each continent over the period 1052-1992.

2 Data Manipulation and Visualization

```
#importing dataset

demscore <- read.csv("dem_score.csv", check.names = FALSE)
head(demscore)
```

```
##       country 1952 1957 1962 1967 1972 1977 1982 1987 1992
## 1     Albania   -9   -9   -9   -9   -9   -9   -9   -9    5
## 2   Argentina   -9   -1   -1   -9   -9   -9   -8    8    7
## 3     Armenia   -9   -7   -7   -7   -7   -7   -7   -7    7
## 4   Australia   10   10   10   10   10   10   10   10   10
## 5     Austria   10   10   10   10   10   10   10   10   10
## 6 Azerbaijan   -9   -7   -7   -7   -7   -7   -7   -7    1
```

```
nrow(demscore)
```

```
## [1] 96
```

```
sum(is.na(demscore))
```

```
## [1] 0
```

```
moltendf <- melt(demscore, id = "country", variable.name = "Year")

head(moltendf)
```

```
##       country Year value
## 1     Albania 1952    -9
## 2   Argentina 1952    -9
## 3     Armenia 1952    -9
## 4   Australia 1952    10
## 5     Austria 1952    10
## 6 Azerbaijan 1952    -9
```

```
#renaming value to DemIndex

demscoreNew <- rename_(moltendf, "DemIndex" = "value")
head(demscoreNew)
```

```
##       country Year DemIndex
## 1     Albania 1952       -9
## 2   Argentina 1952       -9
## 3     Armenia 1952       -9
## 4   Australia 1952       10
## 5     Austria 1952       10
## 6 Azerbaijan 1952       -9
```

```
sum(is.na(demscoreNew)) #checking for NA
```

```
## [1] 0
```

```
#create an interval based on DemIndex and create a new column called "interval"

demscoreNew$Interval <- cut(demscoreNew$DemIndex, c(-11,-6,0,5,10),)

head(demscoreNew)
```

```
##       country Year DemIndex Interval
```

```
## 1    Albania 1952       -9 (-11,-6]
## 2  Argentina 1952       -9 (-11,-6]
## 3    Armenia 1952       -9 (-11,-6]
## 4  Australia 1952       10   (5,10]
## 5    Austria 1952       10   (5,10]
## 6 Azerbaijan 1952       -9 (-11,-6]
```

```
# taking the count grouping by year and interval to plot a stacked bar plot to show the democracy index

demCount<- demscoreNew%>%
  group_by(Year,Interval)%>%
  summarise(c= length(country))

head(demCount)
```

```
## # A tibble: 6 x 3
## # Groups:   Year [2]
##   Year  Interval     c
##   <fct> <fct>    <int>
## 1 1952  (-11,-6]    48
## 2 1952  (-6,0]      14
## 3 1952  (0,5]        9
## 4 1952  (5,10]      25
## 5 1957  (-11,-6]    48
## 6 1957  (-6,0]      11
```

```
a<-levels(demCount$Interval)
str(a)
```

```
##  chr [1:4] "(-11,-6]" "(-6,0]" "(0,5]" "(5,10]"
```

```
#stacked bar plot

ggplot(demCount, aes(fill= Interval,x= Year, y= c))+
  geom_bar(position = "stack",stat = "identity", width = 0.5)+
   geom_text(aes(label=paste0(c,"%")),size = 3, position = position_stack(vjust = 0.5),colour= "darksla
  labs(title="Democracy Index level distribution over the years",caption="Figure1:Democracy Index distri
y="Count of countries", x="Year")+
    scale_fill_brewer(palette = "Spectral",name= "DemocracyIndex Range", breaks = levels(demCount$Interv

theme(panel.grid.minor = element_blank())
```

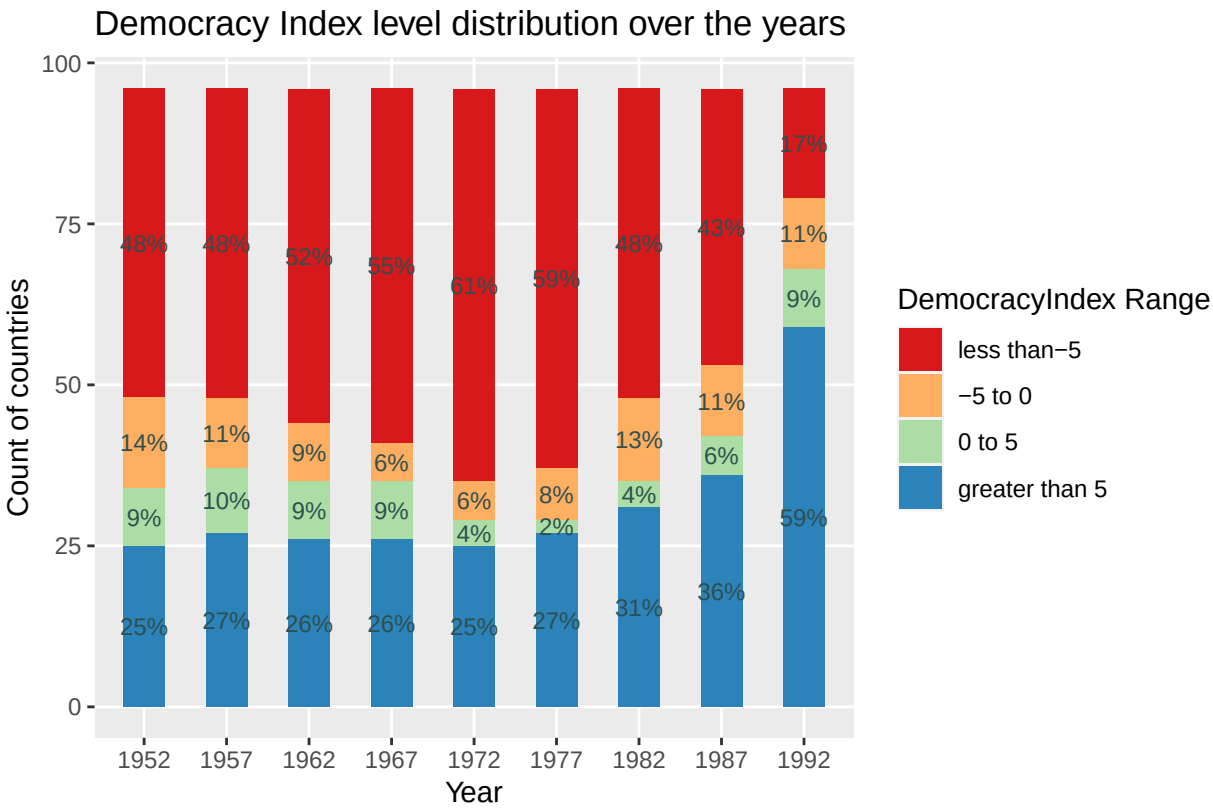


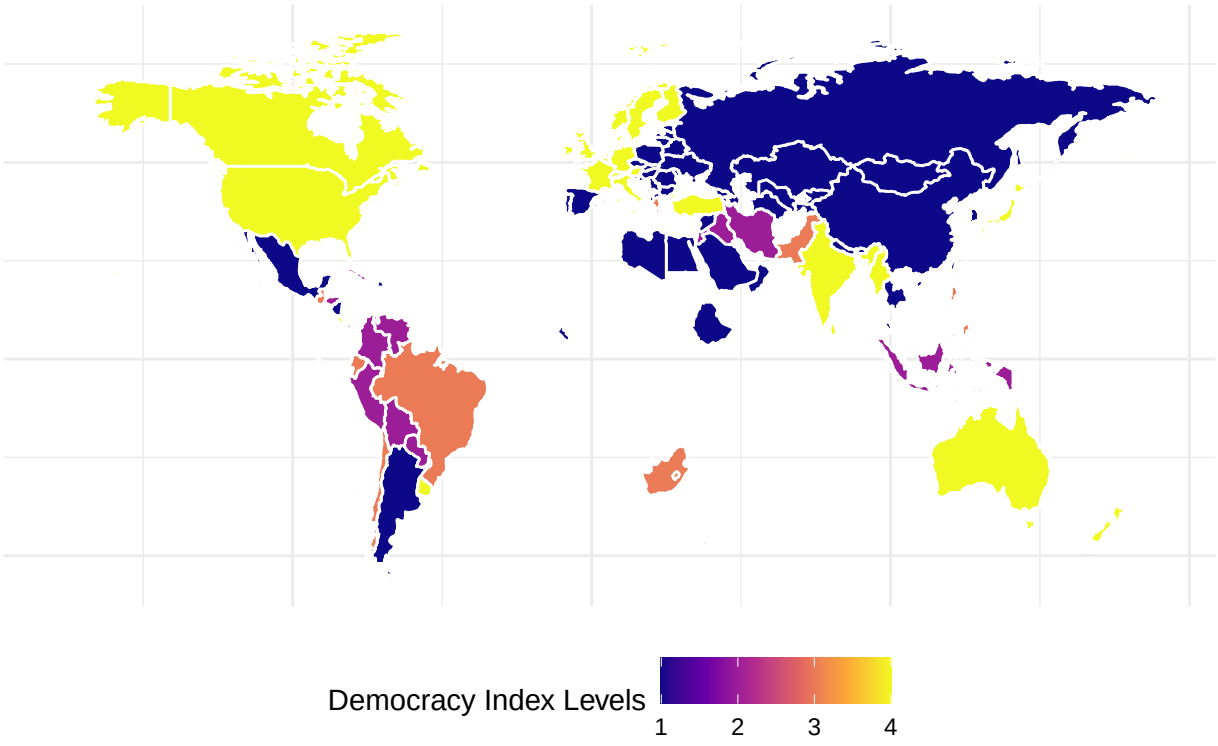
Figure1:Democracy Index distribution:1952−1992

```r
#Because of the name mismatch of counrties between world_map, ccoodes(), and dem_score data frames, con


demscoreNew$ISO2<-countrycode(sourcevar = demscoreNew[, "country"],
                              origin = "country.name",
                              destination = "iso2c")
head(demscoreNew)
```

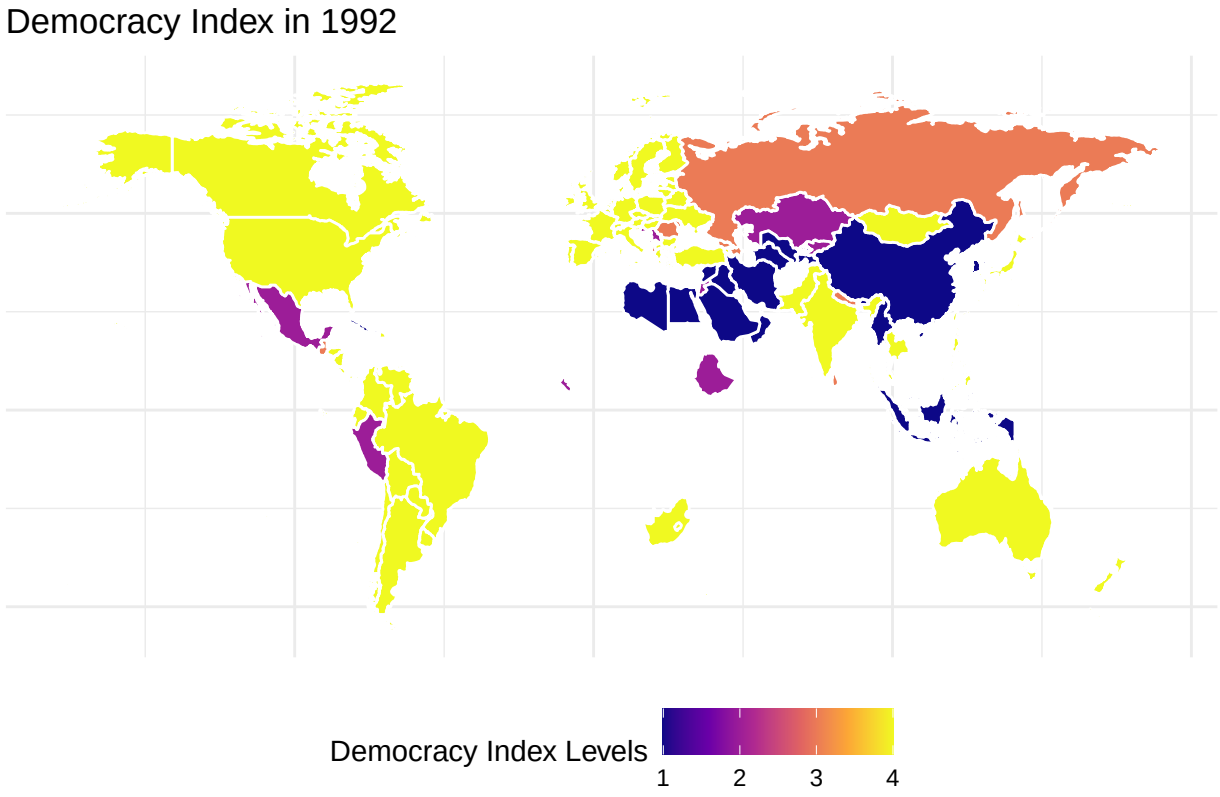
```
##      country Year DemIndex Interval ISO2
## 1    Albania 1952       -9 (-11,-6]   AL
## 2  Argentina 1952       -9 (-11,-6]   AR
## 3    Armenia 1952       -9 (-11,-6]   AM
## 4  Australia 1952       10   (5,10]   AU
## 5    Austria 1952       10   (5,10]   AT
## 6 Azerbaijan 1952       -9 (-11,-6]   AZ
```

```r
#creating function to plot democracy index levels on a map.

plot_dem_on_map <- function(df, year)
{


demMap<- df%>%
  filter(Year==year)

world_map <- map_data("world")
world_map$ISO2<-countrycode(sourcevar = world_map[, "region"],
                              origin = "country.name",
                              destination = "iso2c", warn = FALSE)
demMap <- left_join(demMap, world_map, by = "ISO2")

ggplot(demMap, aes(long, lat, group = group))+
  geom_polygon(aes(fill = as.numeric(Interval)),color = "white")+
  labs(title=paste("Democracy Index in",year),fill= "Democracy Index Levels", caption = "1- Authoratori
  scale_fill_viridis_c(option = "C")+
  theme_minimal()+
  theme(axis.title = element_blank(),
        axis.text =element_blank(), legend.position="bottom")

}
```

```r
plot_dem_on_map(demscoreNew,1952)
```




```
plot_dem_on_map(demscoreNew,1992)
```




```
#merging demscoreNew with ccodes() to get continent info.


democracyInd_df <- merge(demscoreNew,ccodes()[,c("ISO2","continent")],by.x=,by.y=)
 head(democracyInd_df)
```

```
##   ISO2 country Year DemIndex Interval continent
## 1   AL Albania 1952       -9 (-11,-6]    Europe
## 2   AL Albania 1987       -9 (-11,-6]    Europe
## 3   AL Albania 1992        5    (0,5]    Europe
## 4   AL Albania 1957       -9 (-11,-6]    Europe
## 5   AL Albania 1972       -9 (-11,-6]    Europe
## 6   AL Albania 1962       -9 (-11,-6]    Europe
```

```
# taking average of DemIndex to plot average democracy index of each continent over the period 1952-199

demIndexAverage <- democracyInd_df%>%
  group_by(continent,Year)%>%
summarise(a= mean(DemIndex))

head(demIndexAverage)
```

```
## # A tibble: 6 x 3
## # Groups:   continent [1]
##   continent Year      a
##   <chr>     <fct> <dbl>
## 1 Africa    1952   -5
```

```
## 2 Africa    1957   -5
## 3 Africa    1962   -5
## 4 Africa    1967   -5
## 5 Africa    1972   -5
## 6 Africa    1977   -4.4
```

```
ggplot(demIndexAverage,aes(x= as.factor(Year), y= a, group = continent, colour = continent ))+

geom_line(size =1.5)+
geom_point(aes(x = as.factor(Year), y = a,group = continent, colour = continent), size = 3)+
  scale_colour_brewer(palette = "Set2")+
labs(title = "Average Democracy Index over the years based on continenets",
caption = "Figure3: Average Democracy Index of continenets",
x = "Year", y = "Average Democracy Index")+
theme_minimal()+
  theme(panel.grid.major = element_blank())
```

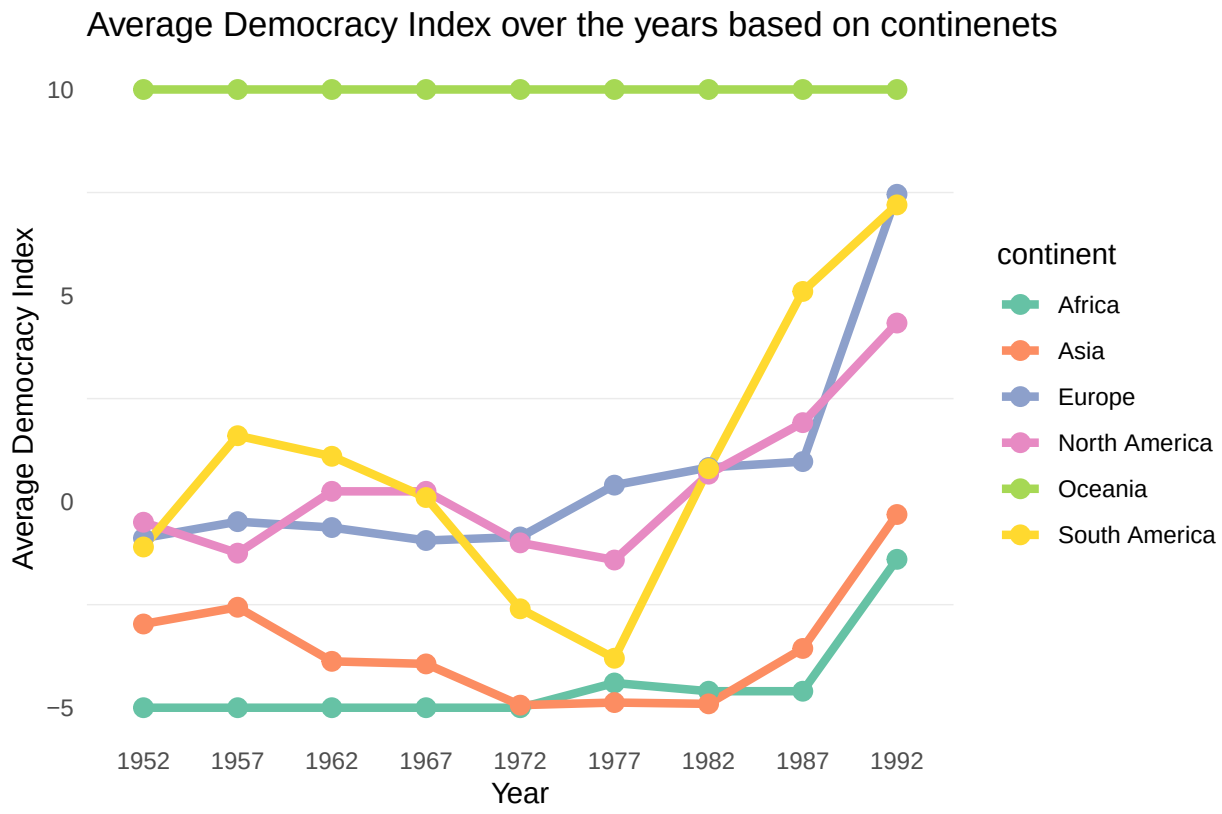


Figure3: Average Democracy Index of continenets

3 Data Analysis using the above Data Visualization

3.1 How did the world change in terms of democracy during the period 1952 to 1992 and which continent went through the biggest shift from autocracy to democracy?

The chart "Democracy Index level distribution over the years "depicts the slow rise in the number of democracies over a period starting from 1952 to 1992. It can be seen that in 1952 there were very few democratic nations over the world, and it accounted for only 25%. At this time almost 50% of the countries were under the authoritarian ruling. Up until 1972, there has been no significant change in the number of democratic regimes, whereas the autocratic leadership around the world rose over 60% by 1972. Countries which were in either Hybrid regime or flawed democracy slid into autocracy run from 1952 to 1972, resulted in the spike of autocratic nations. This pattern was reversed by the start of 1977. It is clear from the plot that there was a substantial increase in democratic regimes from 1977 to 1992. It is evident from the plot that from1972 to 1987 there was a steady increase in the democratic countries, from 25% to 36%. After this period, there was a substantial rise in the number of countries that moved to a democratic system, which accounted for 59% of the countries. Also, a high number of purely autocratic nations moved to hybrid or flawed leadership during this period, it left with dictatorship nations accounting for only 17%.

The world map "Democracy index in 1952" illustrates a detailed view of each country and the change in their regime in 1952 and in 1992. In 1952, most of the countries were in either fully or hybrid autocratic ruling. For instance, take a look at South America, It is clear from the plot that South America was in a highly autocratic leadership and Uruguay was the only country in this continent which had a fully democratic ruling in 1952. Chile, Ecuador, and Brazil were leaning towards democracy this year, whereas the rest of the countries leaned towards an authoritarian system.

The plot "Democracy Index in 1992" comparing with the "Democracy index in 1952" clearly shows the transition of South America from autocratic to a democratic region over the period of 40 years. There was only one country in this continent that remained in the same regime as in 1952, which was Peru. Peru was under a hybrid system throughout this period.Furthermore, comparing the two plots, it is clear that a trend towards democracy was present not only in South America but also in other parts of the world as well.

Furthermore, the line chart highlights South America's transition by showing the average democracy index over the period. From 1972 to 1997 south America went from being one of the worst democratic continents to one of the best. In addition, North America and Europe showed similar trends until 1987. However, while the former showed a slight increase in the average democracy index from 1987 to 1992, the latter recorded a significant peak during this period. Asia and Africa did not show any increase in the democracy index until 1982. The continent of Oceania stayed constant during the entire period, having the highest democracy index compared to all other continents.

In conclusion, there was a widespread adoption of democracy around the world from 1952 to1992. Although most continents of the world improved in terms of democracy, South America witnessed the most significant change.