

# STAT 5014 HW 8

David Edwards

11/8/2019

## Problem 3

Read in the data

```
#Run all of this just once.
#Read in Data
Ed.data <- read.csv("./Ed_Data.csv", header = T)

#Remove columns that are not needed
Ed.data <- Ed.data[,c(1, 4, 5:69)]

#Determine if any columns are all NA's
colmunRemove <- c()
for(i in 1:67){
  if(all(is.na(Ed.data[,i]))){
    colmunRemove <- c(colmunRemove, i)
  }
} #Turns out all the columns contain at least 1 value

#Find all rows that contain nothing but NAs
rowRemove <- c()
for(j in 1:nrow(Ed.data)){
  if(all(is.na(Ed.data[j,3:47]))){
    rowRemove <- c(rowRemove, j)
  }
  if(j %% 100 == 0){
    print(j)
  }
}

#Remove the rows
Ed.data <- Ed.data[-rowRemove, ]

#Gather the data
CleanData <- gather(Ed.data, "Year", "Value", -c(i..Country.Name, Indicator.Code), na.rm = F)

#Again remove rows that contain NAs
CleanData_Rm <- CleanData[!is.na(CleanData$Value),]
save.image(file = "CleanData.RData")

load(file = "./CleanData.RData")
names(CleanData_Rm) <- c("Country.Name", "Indicator.Code", "Year", "Value")

#Pick to countries
Iceland <- CleanData_Rm[CleanData_Rm$Country.Name == "Iceland",]
Italy <- CleanData_Rm[CleanData_Rm$Country.Name == "Italy",]
```

```

#Combine there data
twoContries <- rbind(Iceland, Italy)

#Make sure that we compare the same values for each country
twoContries <- spread(twoContries, Indicator.Code, Value)

#I decided to work with 250 variables to make sure that I have enough data and to show that the ideas w
#Since I don't know what any of this data really means I figured which variables I pick doesn't really
twoContries <- twoContries[,1:250]

#Remove the rows that don't contain any information
rowRemove <- c()
for(j in 1:nrow(twoContries)){
  if(all(is.na(twoContries[j,3:250]))){
    rowRemove <- c(rowRemove, j)
  }
}

twoContries <- twoContries[-rowRemove,]

Iceland <- twoContries[1:9,]
Iceland <- select(Iceland, -c(Country.Name,Year))
Italy <- twoContries[10:18,]
Italy <- select(Italy, -c(Country.Name, Year))

Iceland_Avg <- sapply(Iceland, mean)
Italy_Avg <- sapply(Italy, mean)

two_Avg <- rbind(Iceland_Avg, Italy_Avg)

#two_Avg <- two_Avg[,1:250]

kable(t(two_Avg), "latex", longtable = T, booktabs = T, caption = "Average Values")

```

Table 1: Average Values

	Iceland_Avg	Italy_Avg
BAR.NOED.1519.FE.ZS	1.3966667	5.888889e-01
BAR.NOED.1519.ZS	1.7133333	6.744444e-01
BAR.NOED.15UP.FE.ZS	0.9944444	8.752222e+00
BAR.NOED.15UP.ZS	1.2744444	6.945556e+00
BAR.NOED.2024.FE.ZS	1.6133333	1.190000e+00
BAR.NOED.2024.ZS	1.7566667	1.184444e+00
BAR.NOED.2529.FE.ZS	1.3266667	1.566667e+00
BAR.NOED.2529.ZS	1.5511111	1.450000e+00
BAR.NOED.25UP.FE.ZS	0.8655556	1.023333e+01
BAR.NOED.25UP.ZS	1.1222222	8.176667e+00
BAR.NOED.3034.FE.ZS	1.0133333	1.916667e+00
BAR.NOED.3034.ZS	1.3233333	1.713333e+00
BAR.NOED.3539.FE.ZS	1.0766667	2.791111e+00
BAR.NOED.3539.ZS	1.2833333	2.394444e+00
BAR.NOED.4044.FE.ZS	0.7844444	4.010000e+00

BAR.NOED.4044.ZS	1.1322222	3.341111e+00
BAR.NOED.4549.FE.ZS	0.6155556	5.997778e+00
BAR.NOED.4549.ZS	1.0333333	4.903333e+00
BAR.NOED.5054.FE.ZS	0.5033333	8.037778e+00
BAR.NOED.5054.ZS	0.8900000	6.614444e+00
BAR.NOED.5559.FE.ZS	0.5900000	1.106667e+01
BAR.NOED.5559.ZS	0.6600000	9.027778e+00
BAR.NOED.6064.FE.ZS	0.6377778	1.233889e+01
BAR.NOED.6064.ZS	0.6522222	1.020556e+01
BAR.NOED.6569.FE.ZS	0.7500000	1.895444e+01
BAR.NOED.6569.ZS	0.7688889	1.587667e+01
BAR.NOED.7074.FE.ZS	0.8300000	2.088222e+01
BAR.NOED.7074.ZS	0.8500000	1.781778e+01
BAR.NOED.75UP.FE.ZS	1.1944444	2.646778e+01
BAR.NOED.75UP.ZS	1.2455556	2.320778e+01
BAR.POP.1519	21.6666667	3.756444e+03
BAR.POP.1519.FE	10.6666667	1.840222e+03
BAR.POP.15UP	190.5555556	4.645089e+04
BAR.POP.15UP.FE	95.0000000	2.413489e+04
BAR.POP.2024	20.7777778	3.894444e+03
BAR.POP.2024.FE	10.1111111	1.915667e+03
BAR.POP.2529	19.7777778	3.990333e+03
BAR.POP.2529.FE	9.6666667	1.975444e+03
BAR.POP.25UP	148.1111111	3.880000e+04
BAR.POP.25UP.FE	74.2222222	2.037900e+04
BAR.POP.3034	18.2222222	4.074222e+03
BAR.POP.3034.FE	9.0000000	2.028000e+03
BAR.POP.3539	17.0000000	4.059556e+03
BAR.POP.3539.FE	8.3333333	2.030222e+03
BAR.POP.4044	15.8888889	3.967556e+03
BAR.POP.4044.FE	7.8888889	1.994667e+03
BAR.POP.4549	14.8888889	3.815000e+03
BAR.POP.4549.FE	7.2222222	1.933000e+03
BAR.POP.5054	13.3333333	3.532000e+03
BAR.POP.5054.FE	6.5555556	1.807222e+03
BAR.POP.5559	11.5555556	3.351000e+03
BAR.POP.5559.FE	5.6666667	1.737000e+03
BAR.POP.6064	9.8888889	3.144000e+03
BAR.POP.6064.FE	5.0000000	1.660778e+03
BAR.POP.6569	8.5555556	2.764111e+03
BAR.POP.6569.FE	4.4444444	1.503444e+03
BAR.POP.7074	7.0000000	2.360889e+03
BAR.POP.7074.FE	3.7777778	1.341444e+03
BAR.POP.75UP	12.0000000	3.741333e+03
BAR.POP.75UP.FE	6.6666667	2.367778e+03
BAR.PRM.CMPT.1519.FE.ZS	26.3344444	1.468111e+01
BAR.PRM.CMPT.1519.ZS	28.5466667	1.520667e+01
BAR.PRM.CMPT.15UP.FE.ZS	35.2622222	3.389222e+01
BAR.PRM.CMPT.15UP.ZS	37.4055556	3.225556e+01
BAR.PRM.CMPT.2024.FE.ZS	18.5277778	1.497556e+01

BAR.PRM.CMPT.2024.ZS	21.6277778	1.355667e+01
BAR.PRM.CMPT.2529.FE.ZS	21.4244444	2.131444e+01
BAR.PRM.CMPT.2529.ZS	24.9677778	1.960222e+01
BAR.PRM.CMPT.25UP.FE.ZS	38.3333333	3.660667e+01
BAR.PRM.CMPT.25UP.ZS	40.3677778	3.512889e+01
BAR.PRM.CMPT.3034.FE.ZS	25.8488889	2.456556e+01
BAR.PRM.CMPT.3034.ZS	29.2900000	2.254889e+01
BAR.PRM.CMPT.3539.FE.ZS	29.2344444	2.688889e+01
BAR.PRM.CMPT.3539.ZS	32.3533333	2.506667e+01
BAR.PRM.CMPT.4044.FE.ZS	33.3133333	2.960111e+01
BAR.PRM.CMPT.4044.ZS	36.4155556	2.825333e+01
BAR.PRM.CMPT.4549.FE.ZS	37.6855556	3.568333e+01
BAR.PRM.CMPT.4549.ZS	40.0055556	3.415333e+01
BAR.PRM.CMPT.5054.FE.ZS	42.7777778	3.710667e+01
BAR.PRM.CMPT.5054.ZS	44.6366667	3.623000e+01
BAR.PRM.CMPT.5559.FE.ZS	46.0733333	4.320111e+01
BAR.PRM.CMPT.5559.ZS	47.9122222	4.224667e+01
BAR.PRM.CMPT.6064.FE.ZS	50.9000000	4.382222e+01
BAR.PRM.CMPT.6064.ZS	52.0455556	4.332778e+01
BAR.PRM.CMPT.6569.FE.ZS	50.2300000	4.468778e+01
BAR.PRM.CMPT.6569.ZS	51.1166667	4.493889e+01
BAR.PRM.CMPT.7074.FE.ZS	52.2533333	4.435667e+01
BAR.PRM.CMPT.7074.ZS	52.8388889	4.448222e+01
BAR.PRM.CMPT.75UP.FE.ZS	54.6433333	4.301000e+01
BAR.PRM.CMPT.75UP.ZS	55.0088889	4.344000e+01
BAR.PRM.ICMP.1519.FE.ZS	31.1144444	1.553667e+01
BAR.PRM.ICMP.1519.ZS	33.2411111	1.603556e+01
BAR.PRM.ICMP.15UP.FE.ZS	47.7988889	4.135778e+01
BAR.PRM.ICMP.15UP.ZS	49.7900000	3.853000e+01
BAR.PRM.ICMP.2024.FE.ZS	23.6233333	1.604889e+01
BAR.PRM.ICMP.2024.ZS	26.9566667	1.447000e+01
BAR.PRM.ICMP.2529.FE.ZS	28.6055556	2.394222e+01
BAR.PRM.ICMP.2529.ZS	32.2888889	2.186111e+01
BAR.PRM.ICMP.25UP.FE.ZS	53.8088889	4.633778e+01
BAR.PRM.ICMP.25UP.ZS	55.6988889	4.340444e+01
BAR.PRM.ICMP.3034.FE.ZS	34.1366667	2.849444e+01
BAR.PRM.ICMP.3034.ZS	37.6611111	2.591667e+01
BAR.PRM.ICMP.3539.FE.ZS	40.2355556	3.290889e+01
BAR.PRM.ICMP.3539.ZS	43.6311111	3.003222e+01
BAR.PRM.ICMP.4044.FE.ZS	46.0577778	3.786333e+01
BAR.PRM.ICMP.4044.ZS	49.3288889	3.500333e+01
BAR.PRM.ICMP.4549.FE.ZS	53.2611111	4.634667e+01
BAR.PRM.ICMP.4549.ZS	56.0622222	4.291667e+01
BAR.PRM.ICMP.5054.FE.ZS	59.6922222	4.935667e+01
BAR.PRM.ICMP.5054.ZS	62.0300000	4.631778e+01
BAR.PRM.ICMP.5559.FE.ZS	65.8200000	5.831556e+01
BAR.PRM.ICMP.5559.ZS	68.0255556	5.528111e+01
BAR.PRM.ICMP.6064.FE.ZS	72.4366667	6.011111e+01
BAR.PRM.ICMP.6064.ZS	73.6633333	5.763111e+01
BAR.PRM.ICMP.6569.FE.ZS	77.9311111	6.222556e+01

BAR.PRM.ICMP.6569.ZS	78.7211111	6.140111e+01
BAR.PRM.ICMP.7074.FE.ZS	81.1066667	6.115000e+01
BAR.PRM.ICMP.7074.ZS	81.4855556	6.060889e+01
BAR.PRM.ICMP.75UP.FE.ZS	84.6088889	5.910444e+01
BAR.PRM.ICMP.75UP.ZS	84.4266667	5.937778e+01
BAR.PRM.SCHL.1519	5.8666667	4.946667e+00
BAR.PRM.SCHL.1519.FE	5.8822222	4.950000e+00
BAR.PRM.SCHL.15UP	5.5633333	4.497778e+00
BAR.PRM.SCHL.15UP.FE	5.5722222	4.377778e+00
BAR.PRM.SCHL.2024	5.7355556	4.921111e+00
BAR.PRM.SCHL.2024.FE	5.7488889	4.915556e+00
BAR.PRM.SCHL.2529	5.6877778	4.873333e+00
BAR.PRM.SCHL.2529.FE	5.7044444	4.857778e+00
BAR.PRM.SCHL.25UP	5.4711111	4.386667e+00
BAR.PRM.SCHL.25UP.FE	5.4844444	4.245556e+00
BAR.PRM.SCHL.3034	5.6711111	4.832222e+00
BAR.PRM.SCHL.3034.FE	5.6888889	4.804444e+00
BAR.PRM.SCHL.3539	5.5844444	4.756667e+00
BAR.PRM.SCHL.3539.FE	5.6055556	4.712222e+00
BAR.PRM.SCHL.4044	5.5455556	4.663333e+00
BAR.PRM.SCHL.4044.FE	5.5711111	4.593333e+00
BAR.PRM.SCHL.4549	5.4566667	4.535556e+00
BAR.PRM.SCHL.4549.FE	5.4966667	4.434444e+00
BAR.PRM.SCHL.5054	5.4255556	4.416667e+00
BAR.PRM.SCHL.5054.FE	5.4611111	4.292222e+00
BAR.PRM.SCHL.5559	5.3577778	4.224444e+00
BAR.PRM.SCHL.5559.FE	5.3722222	4.070000e+00
BAR.PRM.SCHL.6064	5.3122222	4.133333e+00
BAR.PRM.SCHL.6064.FE	5.3155556	3.978889e+00
BAR.PRM.SCHL.6569	5.1266667	3.796667e+00
BAR.PRM.SCHL.6569.FE	5.1244444	3.616667e+00
BAR.PRM.SCHL.7074	5.0877778	3.707778e+00
BAR.PRM.SCHL.7074.FE	5.0844444	3.537778e+00
BAR.PRM.SCHL.75UP	5.0422222	3.441111e+00
BAR.PRM.SCHL.75UP.FE	5.0288889	3.275556e+00
BAR.SCHL.1519	8.5922222	8.035556e+00
BAR.SCHL.1519.FE	8.7588889	8.138889e+00
BAR.SCHL.15UP	8.8666667	7.690000e+00
BAR.SCHL.15UP.FE	8.9300000	7.334444e+00
BAR.SCHL.2024	11.2800000	1.021222e+01
BAR.SCHL.2024.FE	11.5433333	1.030222e+01
BAR.SCHL.2529	10.7233333	9.754444e+00
BAR.SCHL.2529.FE	10.9866667	9.778889e+00
BAR.SCHL.25UP	8.4844444	7.341111e+00
BAR.SCHL.25UP.FE	8.5111111	6.918889e+00
BAR.SCHL.3034	10.1122222	9.393333e+00
BAR.SCHL.3034.FE	10.3544444	9.333333e+00
BAR.SCHL.3539	9.4711111	8.911111e+00
BAR.SCHL.3539.FE	9.6800000	8.741111e+00
BAR.SCHL.4044	8.9700000	8.462222e+00

BAR.SCHL.4044.FE	9.1300000	8.232222e+00
BAR.SCHL.4549	8.3688889	7.695556e+00
BAR.SCHL.4549.FE	8.4788889	7.328889e+00
BAR.SCHL.5054	7.9255556	7.265556e+00
BAR.SCHL.5054.FE	7.9711111	6.871111e+00
BAR.SCHL.5559	7.4011111	6.356667e+00
BAR.SCHL.5559.FE	7.3766667	5.866667e+00
BAR.SCHL.6064	6.9833333	6.066667e+00
BAR.SCHL.6064.FE	6.9011111	5.602222e+00
BAR.SCHL.6569	6.4388889	5.133333e+00
BAR.SCHL.6569.FE	6.3311111	4.685556e+00
BAR.SCHL.7074	6.2355556	4.985556e+00
BAR.SCHL.7074.FE	6.1011111	4.560000e+00
BAR.SCHL.75UP	5.9688889	4.482222e+00
BAR.SCHL.75UP.FE	5.8111111	4.092222e+00
BAR.SEC.CMPT.1519.FE.ZS	20.9766667	1.314000e+01
BAR.SEC.CMPT.1519.ZS	19.0000000	1.151333e+01
BAR.SEC.CMPT.15UP.FE.ZS	21.3211111	1.987667e+01
BAR.SEC.CMPT.15UP.ZS	20.0544444	2.097444e+01
BAR.SEC.CMPT.2024.FE.ZS	24.3600000	4.752111e+01
BAR.SEC.CMPT.2024.ZS	22.6722222	4.616889e+01
BAR.SEC.CMPT.2529.FE.ZS	26.4300000	3.582778e+01
BAR.SEC.CMPT.2529.ZS	24.8711111	3.590889e+01
BAR.SEC.CMPT.25UP.FE.ZS	19.5711111	1.749667e+01
BAR.SEC.CMPT.25UP.ZS	18.2177778	1.893556e+01
BAR.SEC.CMPT.3034.FE.ZS	27.8811111	3.080667e+01
BAR.SEC.CMPT.3034.ZS	26.1900000	3.108333e+01
BAR.SEC.CMPT.3539.FE.ZS	26.1266667	2.697556e+01
BAR.SEC.CMPT.3539.ZS	24.7566667	2.741000e+01
BAR.SEC.CMPT.4044.FE.ZS	25.5277778	2.381667e+01
BAR.SEC.CMPT.4044.ZS	23.3300000	2.448000e+01
BAR.SEC.CMPT.4549.FE.ZS	22.1988889	1.851778e+01
BAR.SEC.CMPT.4549.ZS	19.8688889	1.981222e+01
BAR.SEC.CMPT.5054.FE.ZS	19.9611111	1.650667e+01
BAR.SEC.CMPT.5054.ZS	17.4222222	1.768667e+01
BAR.SEC.CMPT.5559.FE.ZS	16.2355556	1.095778e+01
BAR.SEC.CMPT.5559.ZS	14.1755556	1.247889e+01
BAR.SEC.CMPT.6064.FE.ZS	13.1922222	9.934444e+00
BAR.SEC.CMPT.6064.ZS	11.7022222	1.128222e+01
BAR.SEC.CMPT.6569.FE.ZS	9.3555556	6.571111e+00
BAR.SEC.CMPT.6569.ZS	8.3755556	7.526667e+00
BAR.SEC.CMPT.7074.FE.ZS	7.9444444	6.325556e+00
BAR.SEC.CMPT.7074.ZS	7.0088889	7.216667e+00
BAR.SEC.CMPT.75UP.FE.ZS	6.1033333	5.068889e+00
BAR.SEC.CMPT.75UP.ZS	5.3788889	5.762222e+00
BAR.SEC.ICMP.1519.FE.ZS	67.2600000	8.363889e+01
BAR.SEC.ICMP.1519.ZS	64.8100000	8.306333e+01
BAR.SEC.ICMP.15UP.FE.ZS	37.5533333	4.418333e+01
BAR.SEC.ICMP.15UP.ZS	34.4544444	4.840000e+01
BAR.SEC.ICMP.2024.FE.ZS	36.7977778	7.607556e+01

BAR.SEC.ICMP.2024.ZS	34.2911111	7.857556e+01
BAR.SEC.ICMP.2529.FE.ZS	40.9166667	6.375111e+01
BAR.SEC.ICMP.2529.ZS	37.8033333	6.719222e+01
BAR.SEC.ICMP.25UP.FE.ZS	33.0477778	3.743778e+01
BAR.SEC.ICMP.25UP.ZS	29.6688889	4.178556e+01
BAR.SEC.ICMP.3034.FE.ZS	42.9088889	5.873667e+01
BAR.SEC.ICMP.3034.ZS	39.5922222	6.237111e+01
BAR.SEC.ICMP.3539.FE.ZS	41.2455556	5.523667e+01
BAR.SEC.ICMP.3539.ZS	38.1066667	5.854778e+01
BAR.SEC.ICMP.4044.FE.ZS	40.1533333	4.961333e+01
BAR.SEC.ICMP.4044.ZS	35.8500000	5.303444e+01
BAR.SEC.ICMP.4549.FE.ZS	36.3233333	4.112333e+01
BAR.SEC.ICMP.4549.ZS	31.5477778	4.489444e+01
BAR.SEC.ICMP.5054.FE.ZS	32.6433333	3.688111e+01
BAR.SEC.ICMP.5054.ZS	27.6455556	4.047111e+01
BAR.SEC.ICMP.5559.FE.ZS	28.7455556	2.658444e+01
BAR.SEC.ICMP.5559.ZS	23.9366667	3.043333e+01
BAR.SEC.ICMP.6064.FE.ZS	23.4644444	2.396889e+01
BAR.SEC.ICMP.6064.ZS	19.7844444	2.735556e+01
BAR.SEC.ICMP.6569.FE.ZS	18.5966667	1.675556e+01
BAR.SEC.ICMP.6569.ZS	15.5322222	1.941333e+01
BAR.SEC.ICMP.7074.FE.ZS	15.8522222	1.599222e+01
BAR.SEC.ICMP.7074.ZS	13.0577778	1.840778e+01
BAR.SEC.ICMP.75UP.FE.ZS	12.6222222	1.281222e+01
BAR.SEC.ICMP.75UP.ZS	10.4900000	1.463444e+01
BAR.SEC.SCHL.1519	2.7200000	3.086667e+00
BAR.SEC.SCHL.1519.FE	2.8733333	3.183333e+00
BAR.SEC.SCHL.15UP	2.8500000	2.994444e+00
BAR.SEC.SCHL.15UP.FE	2.9344444	2.776667e+00
BAR.SEC.SCHL.2024	4.5244444	5.132222e+00
BAR.SEC.SCHL.2024.FE	4.7366667	5.197778e+00
BAR.SEC.SCHL.2529	4.1133333	4.572222e+00
BAR.SEC.SCHL.2529.FE	4.3233333	4.567778e+00

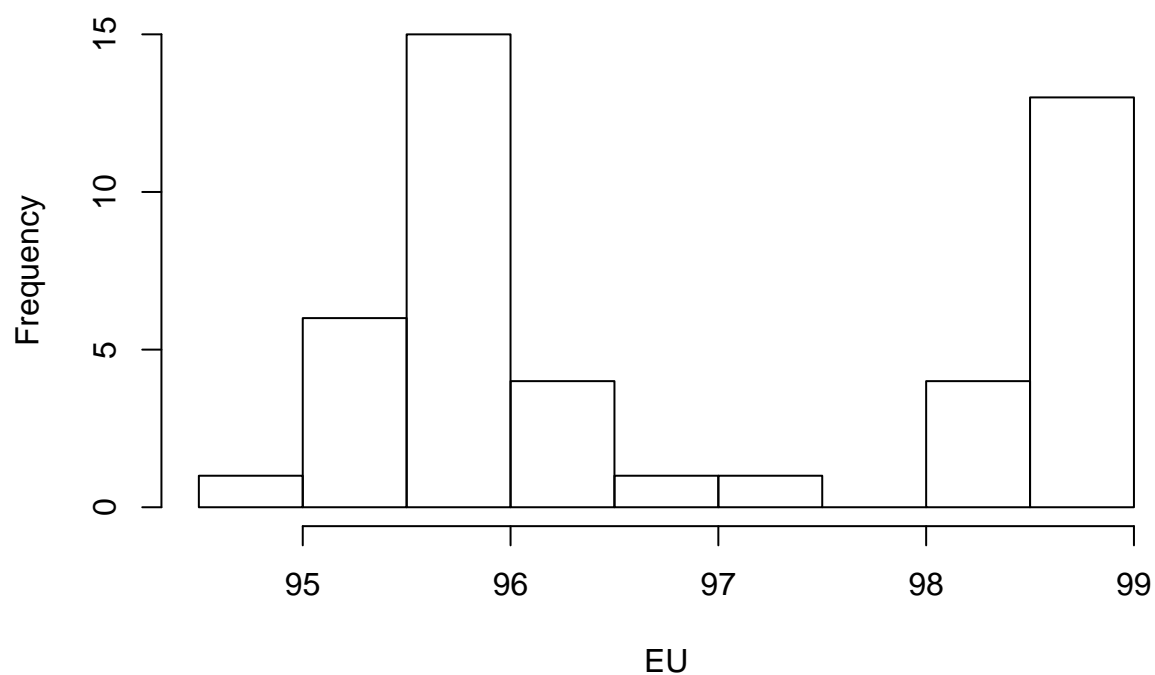
To answer the rest of the question, there were about 880,000 original data points in the data, after cleaning up there were about 5,000,000 data points.

## Problem 4

This is my solution to the scatter plot with marginal histograms using base R. For this problem I just picked an indicator code and compared the values for the Arab and EU regions. Please ignore the first histogram. I just created it to get the data out of it to create the next plot but couldn't figure out how to keep it from displaying.

```
SE.PRM.TENR <- Ed.data[Ed.data$Indicator.Code == "SE.PRM.TENR",]
Arab <- t(SE.PRM.TENR[1,3:47])
EU <- t(SE.PRM.TENR[7,3:47])
H <- hist(EU)
```

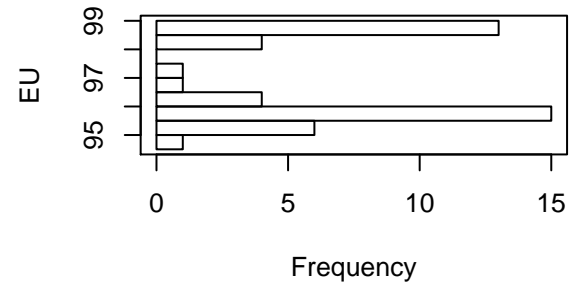
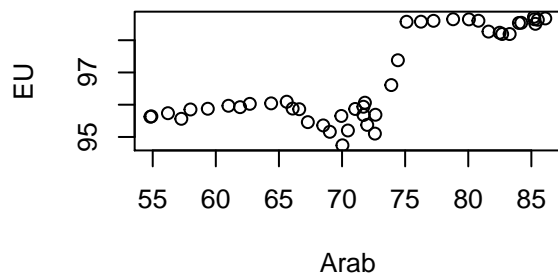
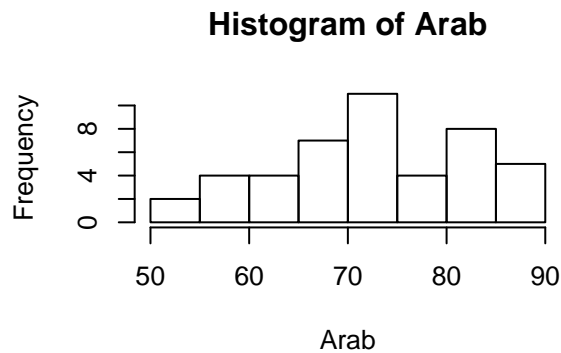
**Histogram of EU**



```
par(mfrow = c(2,2))
par(mfg=c(1,1))
hist(Arab)
par(mfg=c(2,1))
plot(Arab,EU)
```

```
par(mfg=c(2,2))
plot(NULL, type = "n", xlim = c(0, max(H$counts)), ylim = c(range(H$breaks)), xlab = "Frequency", ylab = "EU")
rect(0, H$breaks[1:(length(H$breaks) - 1)], H$counts, H$breaks[2:length(H$breaks)])
```





## Problem 5

Below is my plot using the ggplot2 package. I found this code on the following site: <https://stackoverflow.com/questions/8545035/scatterplot-with-marginal-histograms-in-ggplot2>

```
hist_top <- ggplot()+geom_histogram(aes(Arab))
empty <- ggplot()+geom_point(aes(1,1), colour="white")+
  theme(axis.ticks=element_blank(),
        panel.background=element_blank(),
        axis.text.x=element_blank(), axis.text.y=element_blank(),
        axis.title.x=element_blank(), axis.title.y=element_blank())

scatter <- ggplot()+geom_point(aes(Arab, EU))
hist_right <- ggplot()+geom_histogram(aes(EU))+coord_flip()

grid.arrange(hist_top, empty, scatter, hist_right, ncol=2, nrow=2, widths=c(4, 1), heights=c(1, 4))

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

