

Analysis of the USJudgeRatings data set

Adrien Toulouse

10/30/2019

Analysis of the USJudgeRatings dataset

Introduction and exploration of the dataset

The USJudgeRatings dataset contains lawyers' ratings of state judges in the US Superior Court in 1977.

```
data(USJudgeRatings)
head(USJudgeRatings)
```

```
##           CONT INTG DMNR DILG CFMG DECI PREP FAMI ORAL WRIT PHYS RTEN
## AARONSON,L.H.  5.7  7.9  7.7  7.3  7.1  7.4  7.1  7.1  7.1  7.0  8.3  7.8
## ALEXANDER,J.M.  6.8  8.9  8.8  8.5  7.8  8.1  8.0  8.0  7.8  7.9  8.5  8.7
## ARMENTANO,A.J.  7.2  8.1  7.8  7.8  7.5  7.6  7.5  7.5  7.3  7.4  7.9  7.8
## BERDON,R.I.    6.8  8.8  8.5  8.8  8.3  8.5  8.7  8.7  8.4  8.5  8.8  8.7
## BRACKEN,J.J.   7.3  6.4  4.3  6.5  6.0  6.2  5.7  5.7  5.1  5.3  5.5  4.8
## BURNS,E.B.     6.2  8.8  8.7  8.5  7.9  8.0  8.1  8.0  8.0  8.0  8.6  8.6
```

```
str(USJudgeRatings)
```

```
## 'data.frame':   43 obs. of  12 variables:
## $ CONT: num  5.7 6.8 7.2 6.8 7.3 6.2 10.6 7 7.3 8.2 ...
## $ INTG: num  7.9 8.9 8.1 8.8 6.4 8.8 9 5.9 8.9 7.9 ...
## $ DMNR: num  7.7 8.8 7.8 8.5 4.3 8.7 8.9 4.9 8.9 6.7 ...
## $ DILG: num  7.3 8.5 7.8 8.8 6.5 8.5 8.7 5.1 8.7 8.1 ...
## $ CFMG: num  7.1 7.8 7.5 8.3 6 7.9 8.5 5.4 8.6 7.9 ...
## $ DECI: num  7.4 8.1 7.6 8.5 6.2 8 8.5 5.9 8.5 8 ...
## $ PREP: num  7.1 8 7.5 8.7 5.7 8.1 8.5 4.8 8.4 7.9 ...
## $ FAMI: num  7.1 8 7.5 8.7 5.7 8 8.5 5.1 8.4 8.1 ...
## $ ORAL: num  7.1 7.8 7.3 8.4 5.1 8 8.6 4.7 8.4 7.7 ...
## $ WRIT: num  7 7.9 7.4 8.5 5.3 8 8.4 4.9 8.5 7.8 ...
## $ PHYS: num  8.3 8.5 7.9 8.8 5.5 8.6 9.1 6.8 8.8 8.5 ...
## $ RTEN: num  7.8 8.7 7.8 8.7 4.8 8.6 9 5 8.8 7.9 ...
```

The data is stored in a dataframe. We are provided with 43 observations and 12 quantitative variables.

The variables are : CONT : Number of contacts of lawyer with judge. INTG : Judicial integrity DMNR : Demeanor DILG : Diligence CFMG : Case flow managing DECI : Prompt decisions PREP : Preparation for trial FAMI : Familiarity with law ORAL : Sound oral rulings WRIT : Sound written rulings PHYS : Physical ability RTEN : Worthy of retention

The variables are the ratings received by the judges and seem to range from 0 to 10 (except the variable CONT which is the number of contact of the lawyer with the judge).

The last variable, RTEN, measures if the lawyers think that a judge is worthy staying in the US Superior Cour . Let's measure the correlations between the 11 first variables and the variable RTEN.

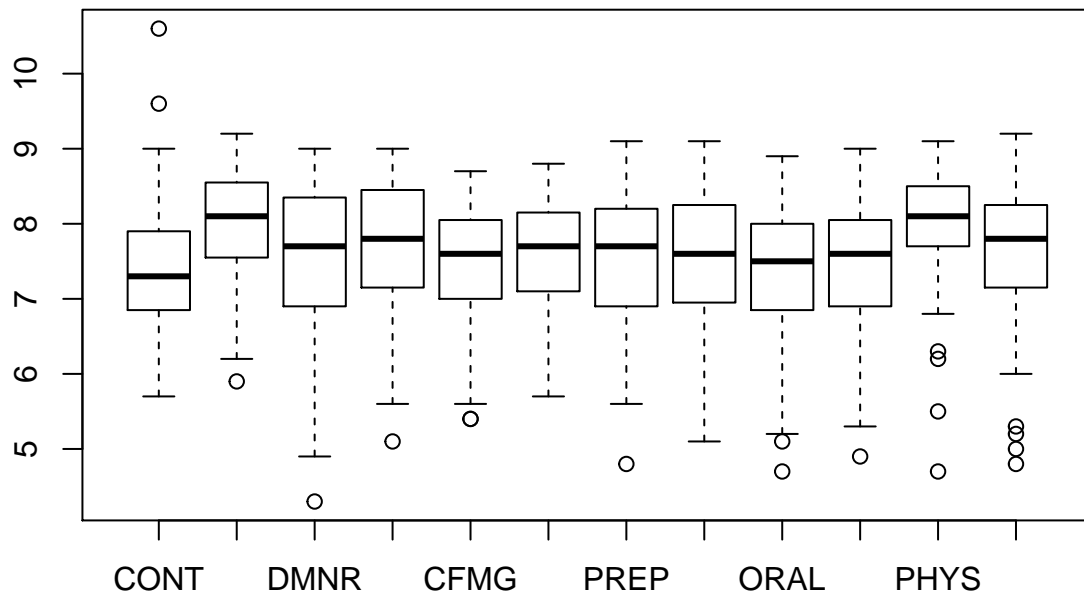
Descriptive statistics analysis of the dataset

```
sum(is.na(USJudgeRatings))
```

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

There are no missing values in the data frame.

```
Outvals = boxplot(USJudgeRatings)
```



We observe the presence of outliers for 10 of the 12 variables (with large values for CONT and with low values for the other variables).

```
max(USJudgeRatings$CONT)
```

```
## [1] 10.6
```

```
rownames(USJudgeRatings)[which.max(USJudgeRatings$CONT)]
```

```
## [1] "CALLAHAN,R.J."
```

```
min(USJudgeRatings$RTEN)
```

```
## [1] 4.8
```

```
rownames(USJudgeRatings)[which.min(USJudgeRatings$RTEN)]
```

```
## [1] "BRACKEN, J. J."
```

The judge with the lowest rating for worthiness of retention is judge Bracken with a rating of 4.8

```
max(USJudgeRatings$RTEN)
```

```
## [1] 9.2
```

```
rownames(USJudgeRatings)[which.max(USJudgeRatings$RTEN)]
```

```
## [1] "RUBINOW, J. E."
```

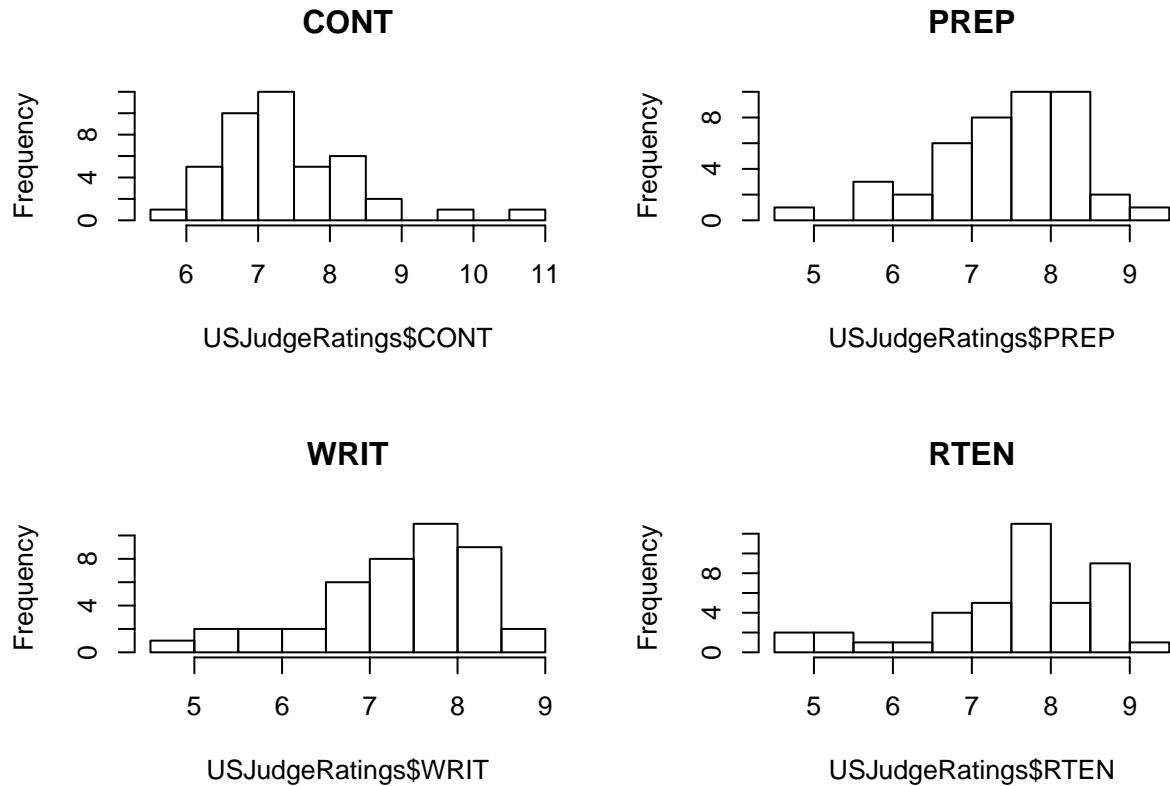
The judge with the highest rating for worthiness of retention is judge Rubinow with a rating of 9.2

```
summary(USJudgeRatings)
```

```
##          CONT          INTG          DMNR          DILG
##  Min.   : 5.700   Min.   :5.900   Min.   :4.300   Min.   :5.100
## 1st Qu.: 6.850   1st Qu.:7.550   1st Qu.:6.900   1st Qu.:7.150
## Median : 7.300   Median :8.100   Median :7.700   Median :7.800
## Mean   : 7.437   Mean   :8.021   Mean   :7.516   Mean   :7.693
## 3rd Qu.: 7.900   3rd Qu.:8.550   3rd Qu.:8.350   3rd Qu.:8.450
## Max.   :10.600   Max.   :9.200   Max.   :9.000   Max.   :9.000
##          CFMG          DECI          PREP          FAMI
##  Min.   :5.400   Min.   :5.700   Min.   :4.800   Min.   :5.100
## 1st Qu.:7.000   1st Qu.:7.100   1st Qu.:6.900   1st Qu.:6.950
## Median :7.600   Median :7.700   Median :7.700   Median :7.600
## Mean   :7.479   Mean   :7.565   Mean   :7.467   Mean   :7.488
## 3rd Qu.:8.050   3rd Qu.:8.150   3rd Qu.:8.200   3rd Qu.:8.250
## Max.   :8.700   Max.   :8.800   Max.   :9.100   Max.   :9.100
##          ORAL          WRIT          PHYS          RTEN
##  Min.   :4.700   Min.   :4.900   Min.   :4.700   Min.   :4.800
## 1st Qu.:6.850   1st Qu.:6.900   1st Qu.:7.700   1st Qu.:7.150
## Median :7.500   Median :7.600   Median :8.100   Median :7.800
## Mean   :7.293   Mean   :7.384   Mean   :7.935   Mean   :7.602
## 3rd Qu.:8.000   3rd Qu.:8.050   3rd Qu.:8.500   3rd Qu.:8.250
## Max.   :8.900   Max.   :9.000   Max.   :9.100   Max.   :9.200
```

We have a symetric distribution for all of the variables since median and mean are always close.

```
par(mfrow=c(2,2))
hist(USJudgeRatings$CONT, main="CONT")
hist(USJudgeRatings$PREP, main="PREP" )
hist(USJudgeRatings$WRIT, main="WRIT")
hist(USJudgeRatings$RTEN, main="RTEN")
```



```
round(sqrt(diag(var(USJudgeRatings))),2)
```

```
## CONT INTG DMNR DILG CFMG DECI PREP FAMI ORAL WRIT PHYS RTEN
## 0.94 0.77 1.14 0.90 0.86 0.80 0.95 0.95 1.01 0.96 0.94 1.10
```

Regarding the dispersion, we look at the interquartile range (given by the boxplots) and the empirical standard deviation. Overall, the dispersions are not very high (around 1). We find that the variables DMNR and RTEN have the largest standard deviation, while the DECI variable has the smallest.

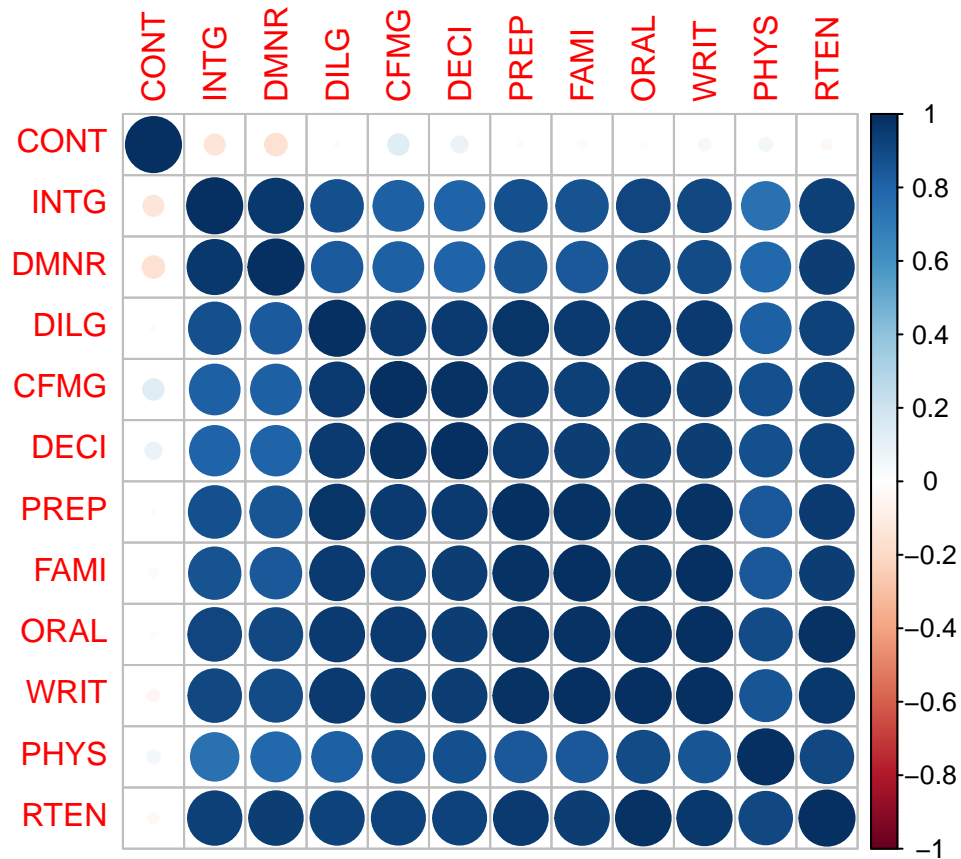
```
round(cor(USJudgeRatings),2)
```

```
##      CONT  INTG  DMNR  DILG  CFMG  DECI  PREP  FAMI  ORAL  WRIT  PHYS  RTEN
## CONT  1.00 -0.13 -0.15  0.01  0.14  0.09  0.01 -0.03 -0.01 -0.04  0.05 -0.03
## INTG -0.13  1.00  0.96  0.87  0.81  0.80  0.88  0.87  0.91  0.91  0.74  0.94
## DMNR -0.15  0.96  1.00  0.84  0.81  0.80  0.86  0.84  0.91  0.89  0.79  0.94
## DILG  0.01  0.87  0.84  1.00  0.96  0.96  0.98  0.96  0.95  0.96  0.81  0.93
## CFMG  0.14  0.81  0.81  0.96  1.00  0.98  0.96  0.94  0.95  0.94  0.88  0.93
## DECI  0.09  0.80  0.80  0.96  0.98  1.00  0.96  0.94  0.95  0.95  0.87  0.92
## PREP  0.01  0.88  0.86  0.98  0.96  0.96  1.00  0.99  0.98  0.99  0.85  0.95
## FAMI -0.03  0.87  0.84  0.96  0.94  0.94  0.99  1.00  0.98  0.99  0.84  0.94
## ORAL -0.01  0.91  0.91  0.95  0.95  0.95  0.98  0.98  1.00  0.99  0.89  0.98
## WRIT -0.04  0.91  0.89  0.96  0.94  0.95  0.99  0.99  0.99  1.00  0.86  0.97
## PHYS  0.05  0.74  0.79  0.81  0.88  0.87  0.85  0.84  0.89  0.86  1.00  0.91
## RTEN -0.03  0.94  0.94  0.93  0.93  0.92  0.95  0.94  0.98  0.97  0.91  1.00
```

```
library(corrplot)
```

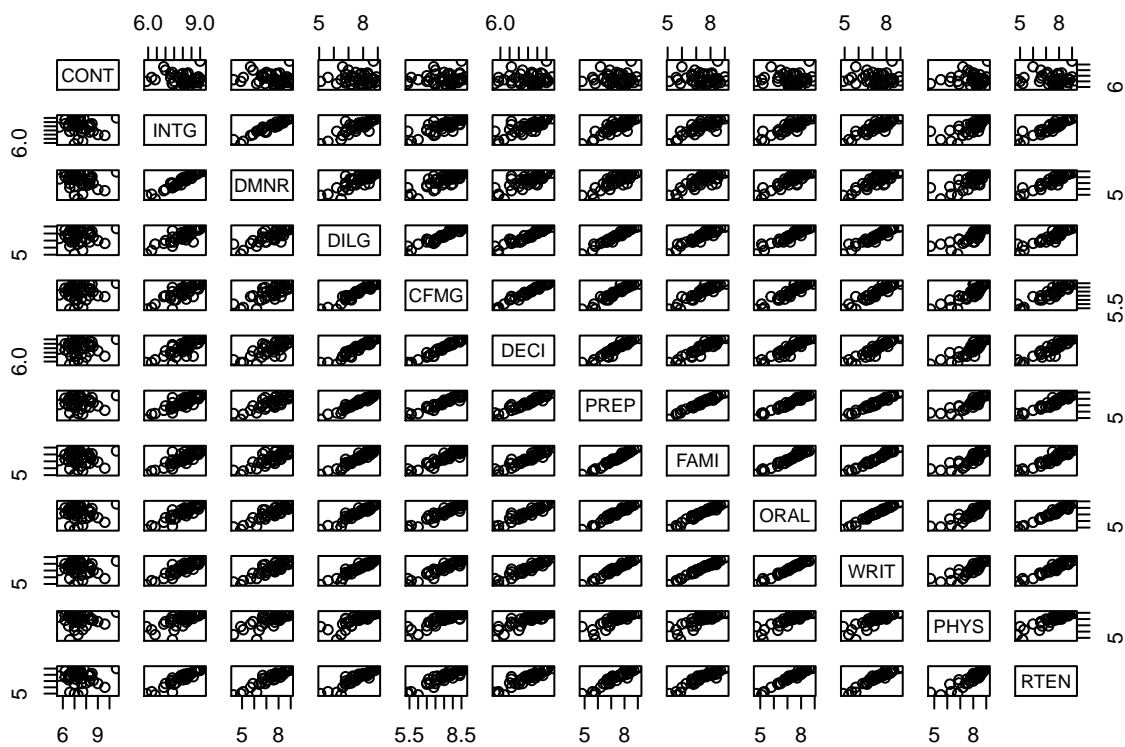
```
## corrplot 0.84 loaded
```

```
corrplot(cor(USJudgeRatings))
```



All the variables have strong positive correlation two by two except the variable CONT which is not correlated to all the other variables. The number of contacts of a lawyer with the judge doesn't seem to explain the ratings received by the judge.

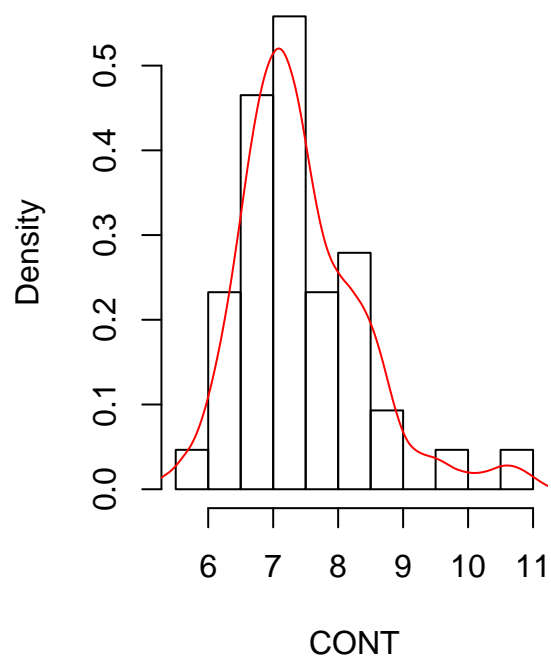
```
pairs(USJudgeRatings)
```



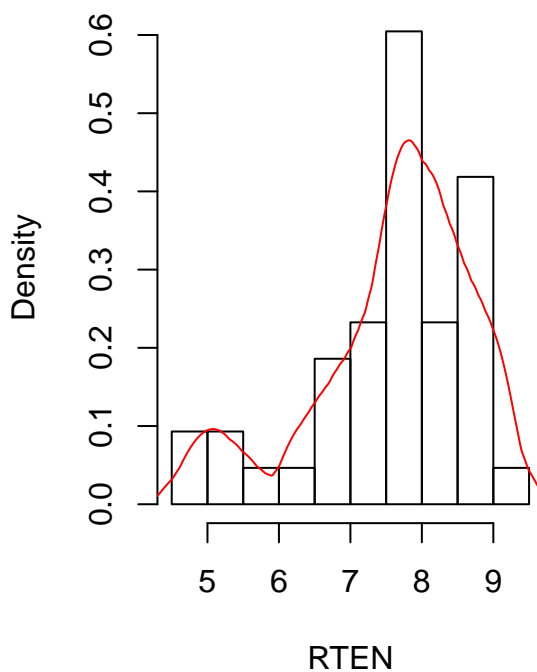
```
par(mfrow=c(1,2))
hist(USJudgeRatings$CONT, probability= TRUE, main="Histogram of CONT", xlab="CONT")
d = density(USJudgeRatings$CONT, kernel = 'c', bw = 0.3)
lines(d, col="red")

hist(USJudgeRatings$RTEN, probability= TRUE, main="Histogram of RTEN" , xlab="RTEN")
d = density(USJudgeRatings$RTEN, kernel = 'o', bw = 0.3)
lines(d, col="red")
```

Histogram of CONT

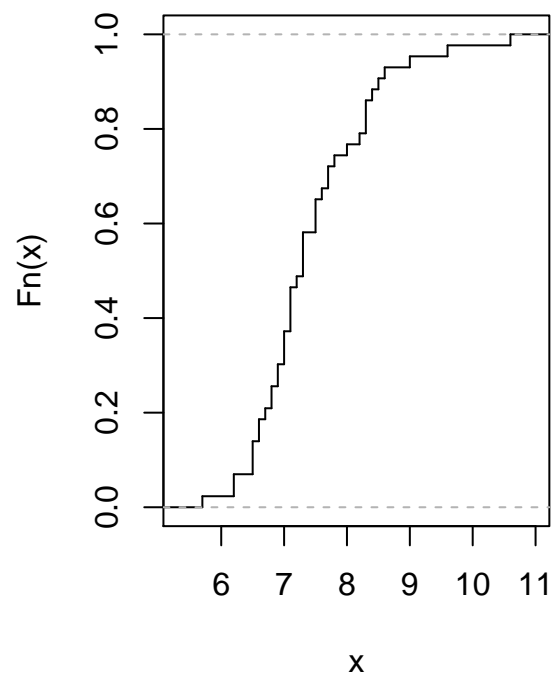


Histogram of RTEN

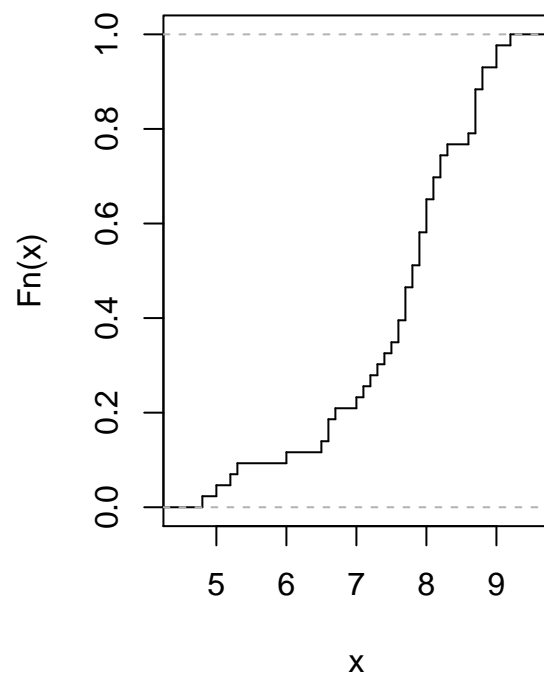


```
par(mfrow=c(1,2))
plot(ecdf(USJudgeRatings$CONT), verticals = TRUE, do.points = FALSE, main = "ECDF CONT")
plot(ecdf(USJudgeRatings$RTEN), verticals = TRUE, do.points = FALSE, main = "ECDF RTEN")
```

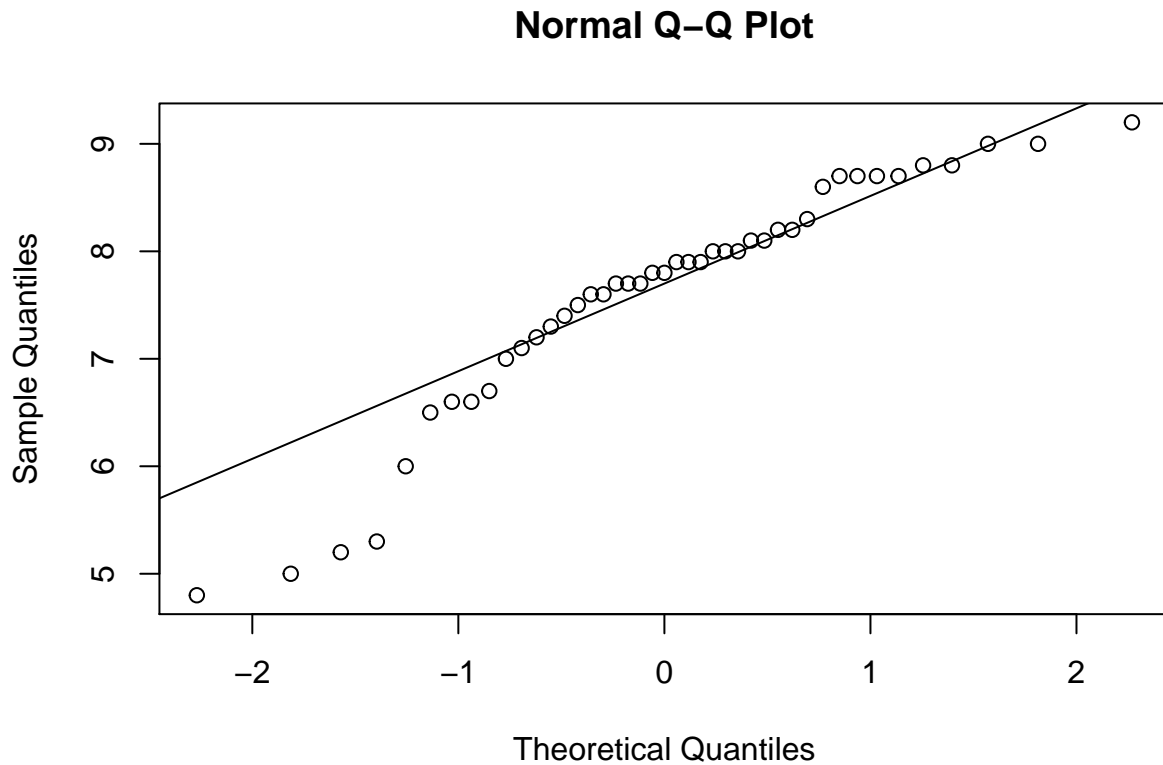
ECDF CONT



ECDF RTEN



```
qqnorm(USJudgeRatings$RTEN)  
qqline(USJudgeRatings$RTEN)
```

The QQ plots suggests that the RTEN variable is Gaussian.

Explaining the RTEN variable with a regression model

We will use RTEN as our dependent variable and try to explain it by fitting a regression model. We will try to find which of the other 11 variables explain the best our dependant variable and therefore which criterion are the most important for lawyers when evaluating if a judge is fit to stay at the Supreme Court.