Data Visualization with R final project

**R code**

africa <- read.csv("african\_crises.csv")

> View(africa)

> dim(africa)

[1] 1059 14

> str(africa)

'data.frame': 1059 obs. of 14 variables:

$ case : int 1 1 1 1 1 1 1 1 1 1 ...

$ cc3 : chr "DZA" "DZA" "DZA" "DZA" ...

$ country : chr "Algeria" "Algeria" "Algeria" "Algeria" ...

$ year : int 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 ...

$ systemic\_crisis : int 1 0 0 0 0 0 0 0 0 0 ...

$ exch\_usd : num 0.0523 0.0528 0.0523 0.0517 0.0513 ...

$ domestic\_debt\_in\_default : int 0 0 0 0 0 0 0 0 0 0 ...

$ sovereign\_external\_debt\_default: int 0 0 0 0 0 0 0 0 0 0 ...

$ gdp\_weighted\_default : num 0 0 0 0 0 0 0 0 0 0 ...

$ inflation\_annual\_cpi : num 3.44 14.15 -3.72 11.2 -3.85 ...

$ independence : int 0 0 0 0 0 0 0 0 0 0 ...

$ currency\_crises : int 0 0 0 0 0 0 0 0 0 0 ...

$ inflation\_crises : int 0 0 0 0 0 0 0 1 0 0 ...

$ banking\_crisis : chr "crisis" "no\_crisis" "no\_crisis" "no\_crisis" ...

> colnames(africa)

[1] "case" "cc3" "country"

[4] "year" "systemic\_crisis" "exch\_usd"

[7] "domestic\_debt\_in\_default" "sovereign\_external\_debt\_default" "gdp\_weighted\_default"

[10] "inflation\_annual\_cpi" "independence" "currency\_crises"

[13] "inflation\_crises" "banking\_crisis"

nrow(africa)

[1] 1059

> ncol(africa)

[1] 14

> attach(africa)

library(dplyr)

> Africa <- africa %>% select(country,year,exch\_usd,domestic\_debt\_in\_default,inflation\_annual\_cpi,banking\_crisis) %>%

+ filter(country %in% c("Egypt","South Africa","Zimbabwe","Algeria","Angola"))

> Africa <- filter(Africa, year > 1880)

> View(Africa)

> str(Africa)

'data.frame': 489 obs. of 6 variables:

$ country : chr "Algeria" "Algeria" "Algeria" "Algeria" ...

$ year : int 1881 1882 1883 1884 1939 1940 1941 1942 1943 1944 ...

$ exch\_usd : num 0.052 0.0508 0.052 0.0522 0.449 ...

$ domestic\_debt\_in\_default: int 0 0 0 0 0 0 0 0 0 0 ...

$ inflation\_annual\_cpi : num 12.62 -12.36 -1.39 -15.94 9.76 ...

$ banking\_crisis : chr "no\_crisis" "no\_crisis" "no\_crisis" "no\_crisis" ...

detach(africa)

> sort(table(Africa$country),decreasing=TRUE)

Egypt South Africa Zimbabwe Angola Algeria

134 114 90 77 74

> Africa$country <- droplevels(Africa$country)

anyNA(Africa)

[1] FALSE

> table(Africa$year)

1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902

2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2

1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924

2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 4 4 4 3

1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946

3 3 4 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5

1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968

5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 4 3 3 3 3 3 4

1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990

4 5 5 5 5 5 5 5 5 5 5 5 4 4 4 4 4 4 4 4 4 4

1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5

2013 2014

5 3

Africa$exch\_usd <- round(Africa$exch\_usd,2)

> Africa$inflation\_annual\_cpi <- round(Africa$inflation\_annual\_cpi,2)

> summary(Africa$exch\_usd)

Min. 1st Qu. Median Mean 3rd Qu. Max.

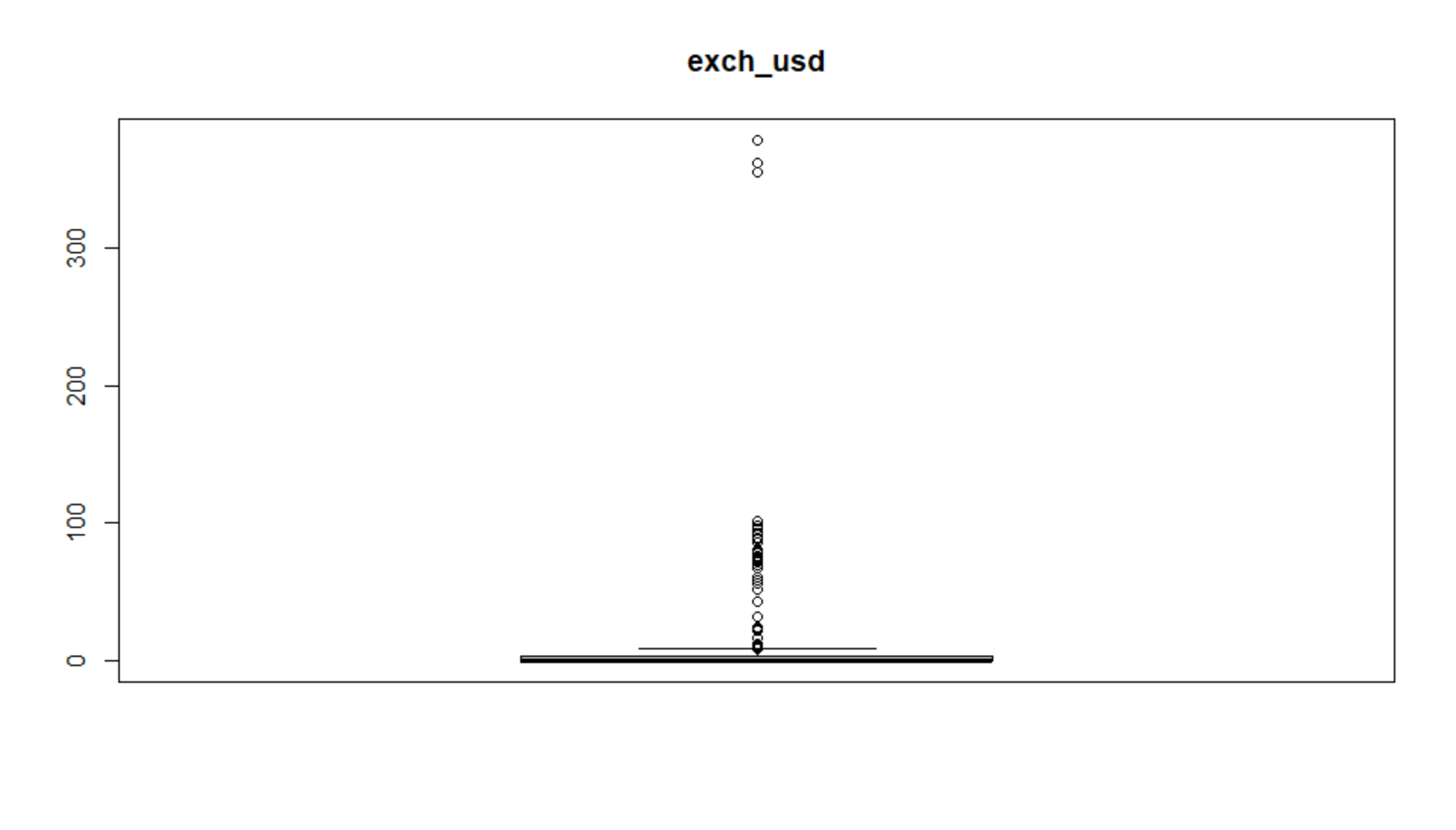
0.00 0.00 0.24 10.31 3.39 378.20

> boxplot(Africa$exch\_usd, main = "exch\_usd")

> summary(Africa$inflation\_annual\_cpi)

Min. 1st Qu. Median Mean 3rd Qu. Max.

-28 1 6 45140 12 21989695



boxplot(Africa$inflation\_annual\_cpi, main = "inflation\_annual\_cpi")

> outlier\_iqr <- function(x){

+ iqr <- IQR(x,na.rm = T,type = 7)

+ q <- quantile(x)

+ upper\_bound = q[4]+(iqr\*1.5)

+ lower\_bound = q[2]-(iqr\*1.5)

+ outliers <- which ((x > upper\_bound) | (x < lower\_bound))

+ return(outliers)

+ }

> print(outlier\_iqr(Africa$inflation\_annual\_cpi))

[1] 4 9 10 11 13 14 52 55 75 76 77 78 79 81 102 123 124 125 126 127 128 129 130 131 132 133

[27] 134 135 136 137 138 139 140 141 186 188 192 213 306 307 400 468 474 475 476 477 478 479 480 481 482 483

[53] 484

> print(outlier\_iqr(Africa$exch\_usd))

[1] 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 137

[27] 138 139 140 141 142 143 144 145 146 147 148 149 150 151 387 388 394 398 399 485 486 487 488 489

|  |
| --- |
| Africa\_data <- Africa[-c(129,130,131,132,133,137,151,479,481,482,483,484,484,485,486,487,488,489),]  > View(Africa\_data)  > str(Africa\_data)  'data.frame': 472 obs. of 6 variables:  $ country : chr "Algeria" "Algeria" "Algeria" "Algeria" ...  $ year : int 1881 1882 1883 1884 1939 1940 1941 1942 1943 1944 ...  $ exch\_usd : num 0.05 0.05 0.05 0.05 0.45 0.49 0.45 0.75 0.5 0.5 ...  $ domestic\_debt\_in\_default: int 0 0 0 0 0 0 0 0 0 0 ...  $ inflation\_annual\_cpi : num 12.62 -12.36 -1.39 -15.94 9.76 ...  $ banking\_crisis : chr "no\_crisis" "no\_crisis" "no\_crisis" "no\_crisis" ... |
|  |
| |  | | --- | | > | |

summary(Africa\_data)

country year exch\_usd domestic\_debt\_in\_default inflation\_annual\_cpi

Length:472 Min. :1881 Min. : 0.000 Min. :0.00000 Min. :-28.500

Class :character 1st Qu.:1935 1st Qu.: 0.000 1st Qu.:0.00000 1st Qu.: 1.155

Mode :character Median :1959 Median : 0.240 Median :0.00000 Median : 5.515

Mean :1958 Mean : 6.581 Mean :0.05508 Mean : 10.767

3rd Qu.:1986 3rd Qu.: 3.333 3rd Qu.:0.00000 3rd Qu.: 12.102

Max. :2014 Max. :97.560 Max. :1.00000 Max. :248.250

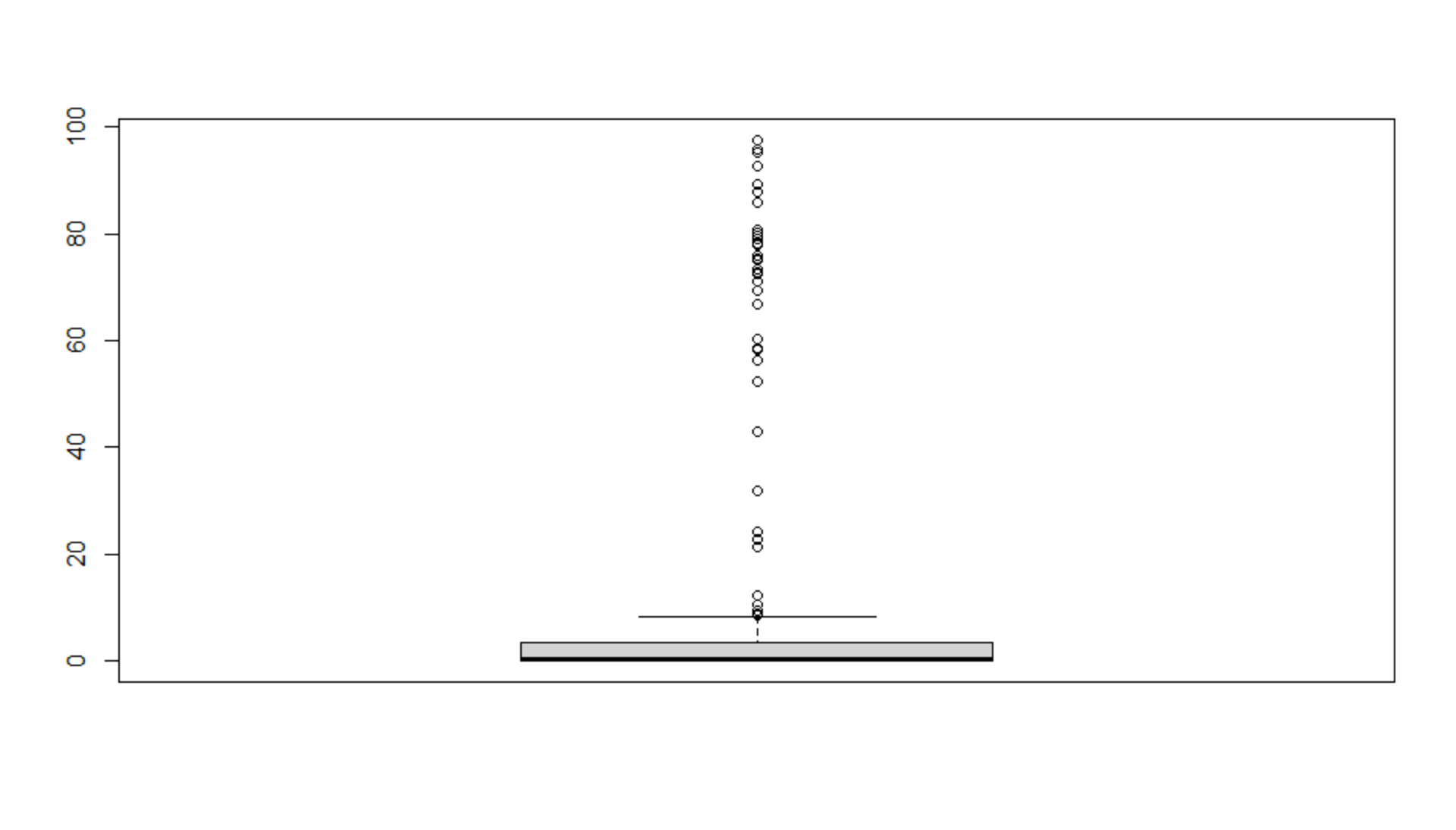
banking\_crisis

Length:472

Class :character

Mode :character

> boxplot(Africa\_data$inflation\_annual\_cpi)

 boxplot(Africa\_data$exch\_usd)

attach(Africa\_data)

> colnames(Africa\_data)

[1] "country" "year" "exch\_usd"

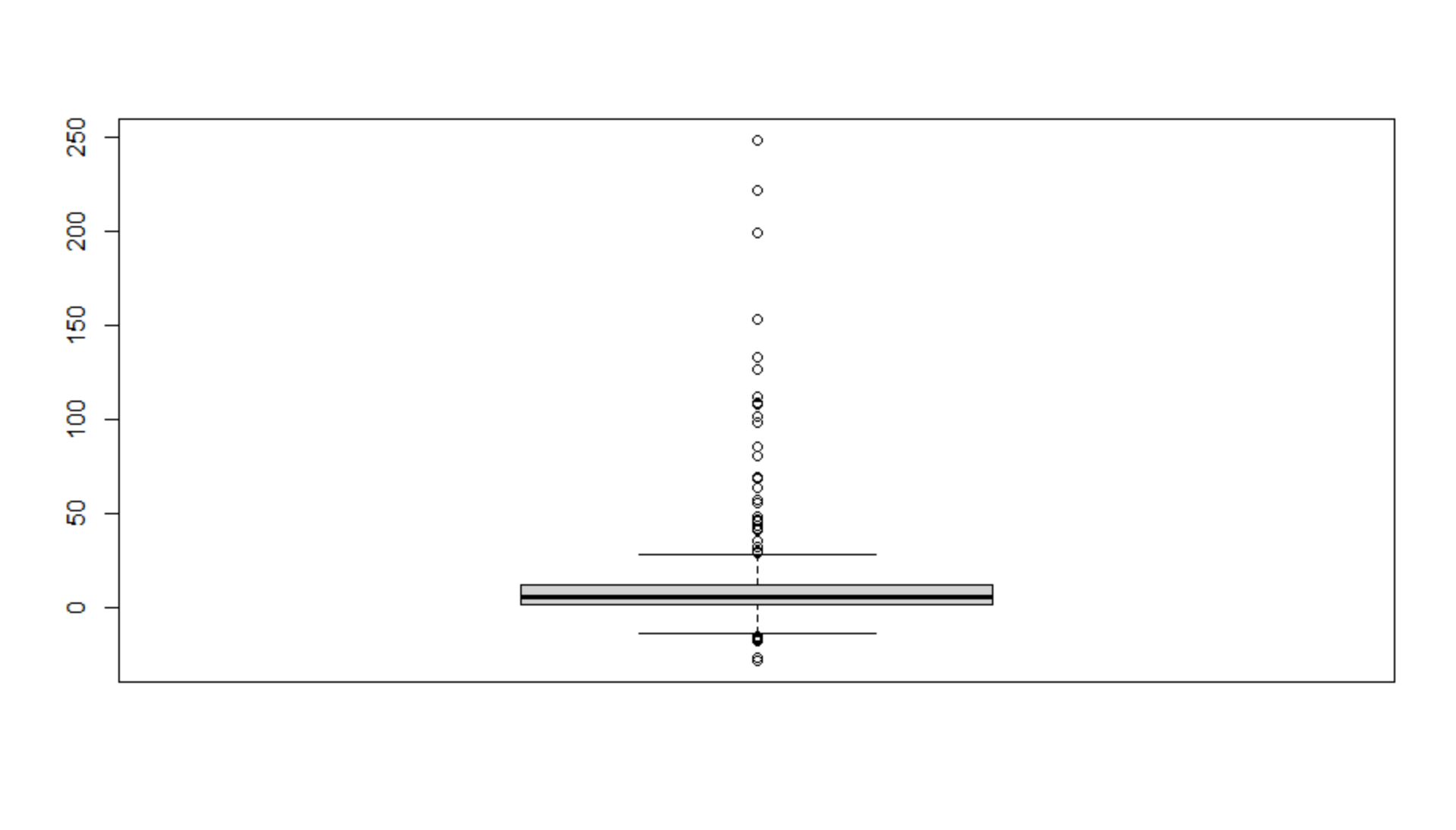
[4] "domestic\_debt\_in\_default" "inflation\_annual\_cpi" "banking\_crisis"

attach(Africa\_data)

> colnames(Africa\_data)

[1] "country" "year" "exch\_usd"

[4] "domestic\_debt\_in\_default" "inflation\_annual\_cpi" "banking\_crisis"



attach(Africa\_data)

> colnames(Africa\_data)

[1] "country" "year" "exch\_usd"

[4] "domestic\_debt\_in\_default" "inflation\_annual\_cpi" "banking\_crisis"

getwd()

[1] "C:/Users/shrut/Desktop/Fall 2020/MET CS 555/project"

> write.csv(Africa\_data,'Africa\_dateset.csv')

> library(aod)

library(stats)

> library(pROC)

tt<-table(banking=Africa\_data$banking\_crisis, debt=Africa\_data$domestic\_debt\_in\_default)

> prop.test(c(22,6),c(446,26),conf.level = 0.95,correct = FALSE)

prop.debt.banking<-tt[1,1]/colSums(tt)[1]

> prop.nodebt.banking<-tt[1,2]/colSums(tt)[2]

risk\_diff <- abs(prop.debt.banking-prop.nodebt.banking)\*100

> model<-glm(as.factor(banking\_crisis)~domestic\_debt\_in\_default, family = binomial)

> summary(model)

Call:

glm(formula = as.factor(banking\_crisis) ~ domestic\_debt\_in\_default,

family = binomial)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.4533 0.3181 0.3181 0.3181 0.7244

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 2.9587 0.2186 13.532 < 2e-16 \*\*\*

domestic\_debt\_in\_default -1.7547 0.5143 -3.412 0.000645 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 212.49 on 471 degrees of freedom

Residual deviance: 203.40 on 470 degrees of freedom

AIC: 207.4

Number of Fisher Scoring iterations: 5

|  |
| --- |
| > exp (cbind (OR = coef (model), confint.default (model)))  OR 2.5 % 97.5 %  (Intercept) 19.272727 12.55534089 29.5840642  domestic\_debt\_in\_default 0.172956 0.06312301 0.4738965 |
|  |
| |  | | --- | | > | |

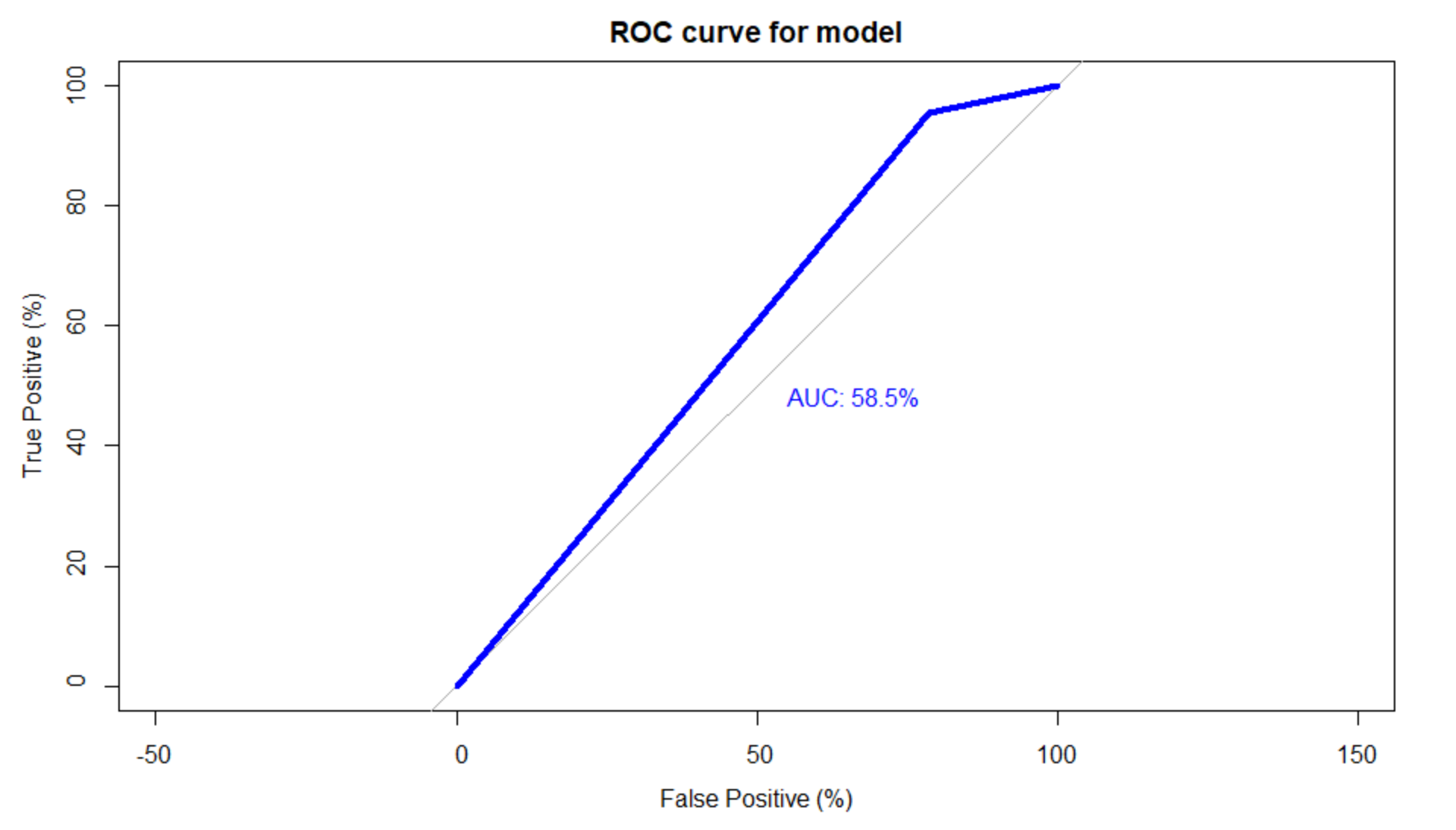
> Africa\_data$prob.model<-predict(model, type=c("response"))

> (g <-roc(Africa\_data$banking\_crisis ~ Africa\_data$prob.model))

> roc(Africa\_data$banking\_crisis ~ Africa\_data$prob.model, plot=TRUE, legacy.axes=T, percent=T,

+ xlab="False Positive (%)", ylab="True Positive (%)", col="blue", lwd=4, print.auc=T, print.auc.x=45, main = "ROC curve for model")

|  |
| --- |
| Call:  roc.formula(formula = Africa\_data$banking\_crisis ~ Africa\_data$prob.model, plot = TRUE, legacy.axes = T, percent = T, xlab = "False Positive (%)", ylab = "True Positive (%)", col = "blue", lwd = 4, print.auc = T, print.auc.x = 45, main = "ROC curve for model")  Data: Africa\_data$prob.model in 28 controls (Africa\_data$banking\_crisis crisis) < 444 cases (Africa\_data$banking\_crisis no\_crisis).  Area under the curve: 58.46% |
|  |
| |  | | --- | | > | |



> model2<-glm(as.factor(banking\_crisis)~Africa\_data$domestic\_debt\_in\_default+Africa\_data$inflation\_annual\_cpi+Africa\_data$exch\_usd,family = binomial)

> summary(model2)

Call:

glm(formula = as.factor(banking\_crisis) ~ Africa\_data$domestic\_debt\_in\_default +

Africa\_data$inflation\_annual\_cpi + Africa\_data$exch\_usd,

family = binomial)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.5541 0.2859 0.3036 0.3281 2.3172

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 3.10222 0.24149 12.846 < 2e-16 \*\*\*

Africa\_data$domestic\_debt\_in\_default -0.23811 0.80458 -0.296 0.767270

Africa\_data$inflation\_annual\_cpi -0.02260 0.00647 -3.493 0.000477 \*\*\*

Africa\_data$exch\_usd 0.02384 0.01986 1.200 0.229979

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 212.49 on 471 degrees of freedom

Residual deviance: 186.88 on 468 degrees of freedom

AIC: 194.88

Number of Fisher Scoring iterations: 6

exp (cbind (OR = coef (model2), confint.default (model2)))

OR 2.5 % 97.5 %

(Intercept) 22.2473683 13.8587740 35.7135050

Africa\_data$domestic\_debt\_in\_default 0.7881124 0.1628264 3.8146227

Africa\_data$inflation\_annual\_cpi 0.9776509 0.9653314 0.9901276

Africa\_data$exch\_usd 1.0241264 0.9850287 1.0647760

Call:

roc.formula(formula = Africa\_data$banking\_crisis ~ Africa\_data$prob.model2, plot = TRUE, legacy.axes = T, percent = T, xlab = "False Positive (%)", ylab = "True Positive (%)", col = "blue", lwd = 2, print.auc = T, print.auc.x = 40, main = "ROC curve for model")

Data: Africa\_data$prob.model2 in 28 controls (Africa\_data$banking\_crisis crisis) < 444 cases (Africa\_data$banking\_crisis no\_crisis).

Area under the curve: 78.66%

> Africa\_data$prob.model2<-predict(model2, type=c("response"))

> (g <-roc(Africa\_data$banking\_crisis ~ Africa\_data$prob.model2))

Call:

roc.formula(formula = Africa\_data$banking\_crisis ~ Africa\_data$prob.model2)

Data: Africa\_data$prob.model2 in 28 controls (Africa\_data$banking\_crisis crisis) < 444 cases (Africa\_data$banking\_crisis no\_crisis).

Area under the curve: 0.7866

> roc(Africa\_data$banking\_crisis ~ Africa\_data$prob.model2, plot=TRUE, legacy.axes=T, percent=T,

+ xlab="False Positive (%)", ylab="True Positive (%)", col="blue", lwd=2, print.auc=T, print.auc.x=40, main = "ROC curve for model")

