#

# CS544 Final R Code with Documentation

# Wilbertiny Jean-Marie  
#

#

#  
Barplot of a normally distribution of the Economic Recording of each country within the dataset.

#

barplot(table(queries$cc3), col = rainbow(13), ylab = "# of recordings",

main = "Economic Recordings by African Countries",

xlim = c(0,20),legend = TRUE)

#

# ... Categorical Data ... country

#

Pie data of categorical data within the dataset. Shows the distribution of the country percentage in terms of how big a crisis is in effect.

#

pie\_data <- table(queries$country)

slice.labels <- names(pie\_data)

slice.percents <- round(pie\_data/sum(pie\_data)\*100)

slice.labels <- paste(slice.labels, slice.percents)

slice.labels <- paste(slice.labels, "%", sep="")

pie(pie\_data, labels = slice.labels,

main = "African Countries with an Economic Crisis")

#

# … Numerical Data … exch\_usd

variable will show the exchange rate out of the full dataset.

Exchange difference between an African country to US Dollar

#

hist(queries$exch\_usd, col = rainbow(9),

main = "Exchange Rate to US Dollars", las = 1,

xlab = "Exchange Rate",

ylab = "# of countries",

xlim = c(0,800),labels = TRUE)

#

# two or more variables ... country & banking crisis

#

Shows the distribution of African countries with a banking crisis. This analysis uses two variables.

#

mycols = c("red4","orange","green","blue",

"pink","magenta","sandybrown","cyan","blue3",

"red","yellow","orange3","gray")

barplot(table(queries$country,queries$banking\_crisis),

col = mycols, ylab = "# of banking crisis",

beside = TRUE, legend = TRUE,

xlim = c(0,65),

main = "African Countries w/ a Banking Crisis")

#

# two or more variables ...

# ... country & domestic debt

#

Shows the distribution of African countries in domestic debt. This analysis uses two variables.

#

barplot(table(queries$country, queries$domestic\_debt\_in\_default),

beside = TRUE, xlim = c(0,60),

names.arg = c("No Debt", "Debt"),

ylab = "# in domestic debt",

col = mycols,legend = TRUE,

main = "African Countries in Domestic Debt")

# normal distribution, defined with two graphs

# numerical data ... year

Shows the distribution of the Years in the dataset. I used two plots to put an emphasis on demonstrate how important the Year variable is to the data that was recorded for this dataset.

#

boxplot(queries$year, horizontal = TRUE, xaxt = "n",

main = "Crisis Distribution by Year",

xlab = "Observations by Year",col = "Pink")

axis(side = 1, at = fivenum(queries$year),

labels = TRUE, las = 2)

text(fivenum(queries$year), rep(1.25,5), srt = 45, adj = 0.75,

labels = c("Min", "Lower Hinge", "Median",

"Upper Hinge", "Max"))

# numerical data w/ distribution ... year

Barplot example of the Year distribution.   
Mean and Standard Deviation are calculated as well.

#

barplot(table(queries$year), xlab = "Year", ylim =c(0,15), col = "Blue3",

ylab = "# of crisis'",

main = "Crisis Distribution by Year")

# mean & sd of the distribution

mean(queries$year)

sd(queries$year)

# Central Limit Theorem

Here I am using the Central Limit Theorem to get a feel of how to collect the normal distribution of the dataset.

#

z <- queries$year

x.sample <- sample(z, size = 1000,

replace = TRUE)

samples <- 1000

sample.size <- 5

xbar <- numeric(samples)

for (i in 1:samples) {

xbar[i] <- mean(sample(z, size = sample.size,

replace = TRUE))

}

par(mfrow = c(2,2))

for (size in c(10, 20, 30, 40)){

for (i in 1:samples) {

xbar[i] <- mean(sample(z, size = size,

replace = TRUE))

}

hist(xbar,

breaks = 15, col = "red3",

main = paste("Sample Size =", size))

cat("Sample Size = ", size, " Mean = ", mean(xbar),

" SD = ", sd(xbar), "\n")

}

mean(queries$year)

#

#

# sampling

#

I used the Simple Random Sample Without Replacement, Systematic Sampling, and Stratified Sampling to conduct my analysis.

#

#srswor

set.seed(123)

library(sampling)

samp\_test <- srswor(25, nrow(queries))

sample.2 <- queries[samp\_test != 0,]

#systematic sampling

N <- nrow(queries)

n <- 25

k <- ceiling(N / n)

r <- sample(k,1)

s <- seq(r, by = k, length = n)

sample.3 <- queries[s,]

#stratified sampling

samp3.year <- queries$year

samp3.Case <- queries$case

data <- data.frame(

Year = samp3.year,

Case = samp3.Case

)

length(table(data))

sample.4 <- strata(data, stratanames = c("Case"),

size = rep(1,2015), method = "srswor",

description = TRUE)

par(mfrow = c(1,4))

barplot(table(queries$case),col = "yellow2")

barplot(table(sample.2$case),col = "yellow2")

barplot(table(sample.3$case),col = "yellow2")

barplot(table(sample.4$Case),col = "yellow2")

mean(queries$case)

sd(queries$case)

mean(sample.2$case)

sd(sample.2$case)

mean(sample.3$case)

sd(sample.3$case)

mean(sample.4$Case)

sd(sample.4$Case)