

## Read data file CSV

lCandidateSpend = read.csv(file='2016\_presidential\_candidate\_expenditures.csv', header=TRUE)

testdf = data.frame(dfCandidateSpend)

colnames(testdf) <- c("Exp\_ID", "Cand\_ID","Cand\_Name","Party","Recipient\_Name",

"Distribution\_Amount","Distribution\_Date", "Recipient\_City","Recipient\_State",

"Recipient\_ZIP","Dist\_Description","Memo\_CD","Memo\_Text","Form\_TP",

"File\_NUM","Transaction\_ID")

## Cleaning up data, first

testdf[4] <- lapply(testdf[4], as.character)

d = subset(testdf, Cand\_Name %in% c("Sanders, Bernard","Lessig, Lawrence",

"Clinton, Hillary Rodham","O'Malley, Martin Joseph","Webb, James Henry Jr."))

d$Party <- 'Democrat'

r = subset(testdf, Cand\_Name %in% c("Rubio, Marco","Paul, Rand","Huckabee, Mike","Graham, Lindsey O.","Christie, Christopher J.",

"Kasich, John R.","Perry, James R. (Rick)","Bush, Jeb","Walker, Scott","Gilmore, James S IIII",

"Carson, Benjamin S.","Santorum, Richard J.","Trump, Donald J.","Cruz, Rafael Edward 'Ted'",

"Fiorina, Carly","Jindal, Bobby"))

r$Party <- 'Republican'

head(r)

g = subset(testdf, Cand\_Name == "Stein, Jill")

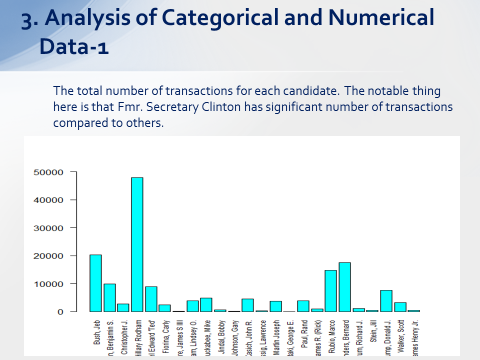
g$Party <- 'Green Party'

l = subset(testdf, Cand\_Name == "Johnson, Gary")

l$Party <- 'Libertarian'

## Combine all 4 dataframes into one, but with party added to the dataframe

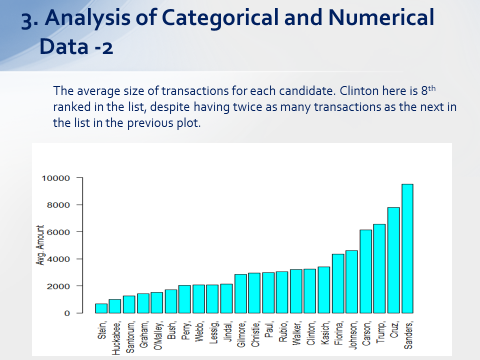
allPartyDf=rbind(d,r,g,l, all=TRUE)



## Creating a barplot for the number of transactions for each candidate

transNumber =table(dplyr::select(allPartyDf,Cand\_Name))

barplot(transNumber,col = "cyan", ylim=c(0,50000),las=2, cex.names=0.75)

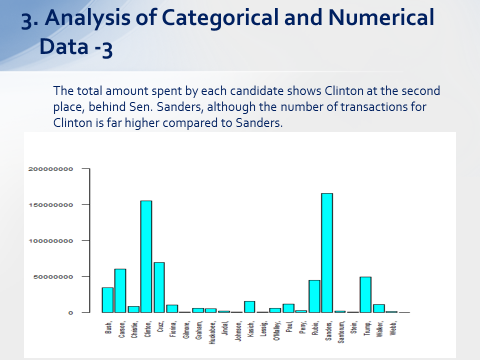


barplot(allPartyDf.avgTrans$avg,

col = "cyan", ylim=c(0,10000),

ylab = "Avg. Amount",las=2,names.arg =gsub( " .\*$", "", allPartyDf.avgTrans$Cand\_Name ))

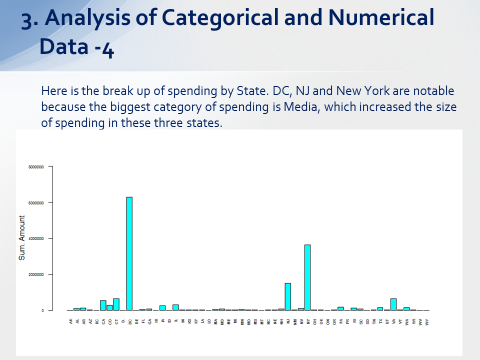
## Senator Sanders' campaign's average transaction amount is more than twice that of Fmr. Secretary Clinton



barplot(allPartyDf2$distSum,

col = "cyan", ylim=c(0,200000000),cex.axis = 0.65,cex.names = 0.65,

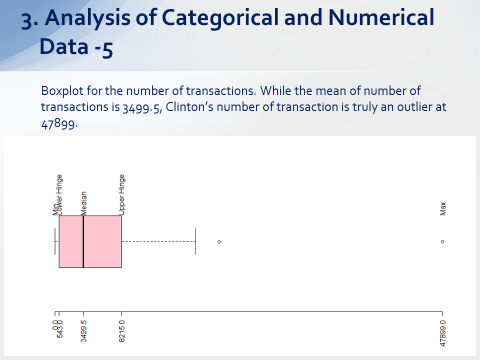
las=2,names.arg =gsub( " .\*$", "", allPartyDf2$Cand\_Name ))



barplot(allPartyDf6$distSum,

col = "cyan", ylim=c(0,80000000),cex.axis = 0.5,cex.names = 0.5,

ylab = "Sum. Amount",las=2,names.arg =gsub( " .\*$", "", allPartyDf6$Recipient\_State))



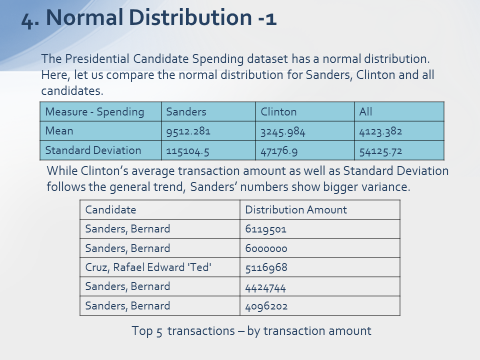
boxplot(transNumber, col=hcl(0), xaxt = "n", horizontal = TRUE)

axis(side = 1, at = fivenum(transNumber), las=2)

text(fivenum(transNumber), rep(1.2,5), srt=90, adj=0,

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

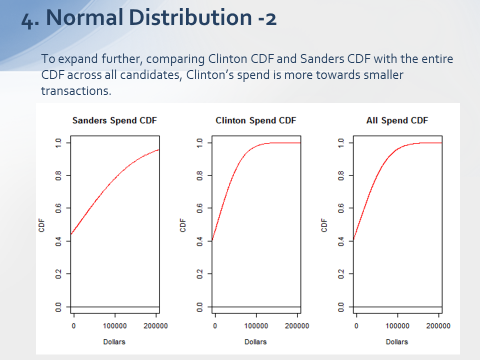
"Upper Hinge", "Max"))



xAll = allPartyDf8 %>% select(Cand\_Name,Distribution\_Amount) %>%

arrange(Distribution\_Amount)

tail(xAll,5)



plot(xClintonNorm, xClintonCdf, type="l", col="red", xlim=c(0,200000), ylim=c(0,1),

xaxt="n", main="Clinton Spend CDF", xlab="Dollars", ylab="CDF")

abline(h=0)

axis(side = 1, at = c(0,100000,200000,300000,400000,500000),

labels = TRUE)

plot(xSandersNorm, xSandersCdf, type="l", col="red", xlim=c(0,200000), ylim=c(0,1),

xaxt="n", main="Sanders Spend CDF", xlab="Dollars", ylab="CDF")

abline(h=0)

axis(side = 1, at = c(0,100000,200000,300000,400000,500000),

labels = TRUE)

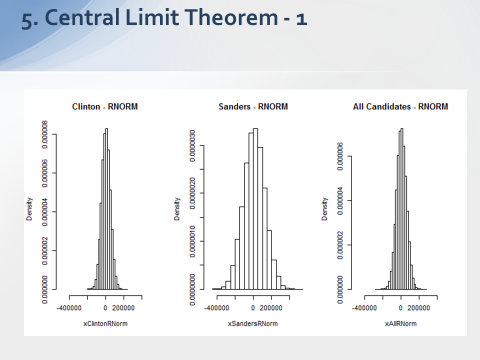
plot(xAllNorm, xAllCdf, type="l", col="red", xlim=c(0,200000), ylim=c(0,1),

xaxt="n", main="All Spend CDF", xlab="Dollars", ylab="CDF")

abline(h=0)

axis(side = 1, at = c(0,100000,200000,300000,400000,500000),

labels = TRUE)



xClintonRNorm = rnorm(xClintonNorm, mean=xClintonMean, sd= xClintonSd)

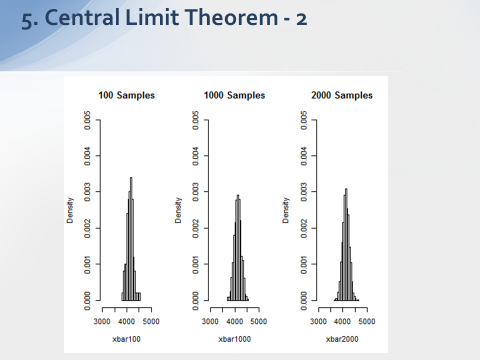
xSandersRNorm = rnorm(xSandersNorm, mean=xSandersMean, sd= xSandersSd)

xAllRNorm = rnorm(xAllNorm, mean=xAllMean, sd= xAllSd)

hist(xClintonRNorm, prob = TRUE, xlim=c(-500000,500000), main = "Clinton - RNORM")

hist(xSandersRNorm, prob = TRUE, xlim=c(-500000,500000), main = "Sanders - RNORM")

hist(xAllRNorm, prob = TRUE, xlim=c(-500000,500000), main = "All Candidates - RNORM")



samples100 <- 100

xbar100 <- numeric(samples)

for (i in 1: samples) {

xbar100[i] <- mean(rnorm(xAllNorm,

mean = xAllMean, sd = xAllSd))

}

hist(xbar100, prob = TRUE,breaks = 15, xlim=c(3000,5000), ylim = c(0, 0.005),main = "100 Samples")

samples1000 <- 1000

xbar1000 <- numeric(samples1000)

for (i in 1: samples1000) {

xbar1000[i] <- mean(rnorm(xAllNorm,

mean = xAllMean, sd = xAllSd))

}

hist(xbar1000, prob = TRUE,breaks = 15, xlim=c(3000,5000), ylim = c(0, 0.005),main = "1000 Samples")

samples2000 <- 2000

xbar2000 <- numeric(samples2000)

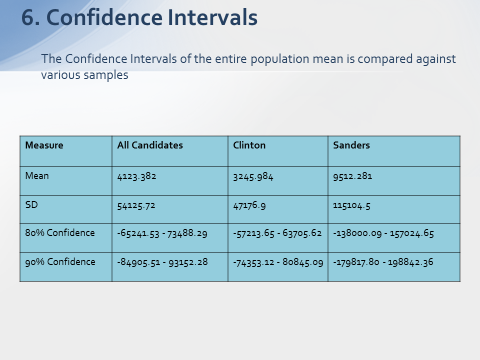
for (i in 1: samples2000) {

xbar2000[i] <- mean(rnorm(xAllNorm,

mean = xAllMean, sd = xAllSd))

}

hist(xbar2000, prob = TRUE,breaks = 15, xlim=c(3000,5000), ylim = c(0, 0.005),main = "2000 Samples")



xbar = xAllMean

xbar

for (i in alpha) {

str <- sprintf("%2d%% Conf Level (alpha = %.2f), CI = %.2f - %.2f",

100\*(1-i), i,

xbar - qnorm(1-i/2) \* xAllSd,

xbar + qnorm(1-i/2) \* xAllSd)

cat(str,"\n")

}

####################################################################

xSandersSd

xbar = xSandersMean

xbar

for (i in alpha) {

str <- sprintf("%2d%% Conf Level (alpha = %.2f), CI = %.2f - %.2f",

100\*(1-i), i,

xbar - qnorm(1-i/2) \* xSandersSd,

xbar + qnorm(1-i/2) \* xSandersSd)

cat(str,"\n")

}

################################################################

xClintonSd

xbar = xClintonMean

xbar

for (i in alpha) {

str <- sprintf("%2d%% Conf Level (alpha = %.2f), CI = %.2f - %.2f",

100\*(1-i), i,

xbar - qnorm(1-i/2) \* xClintonSd,

xbar + qnorm(1-i/2) \* xClintonSd)

cat(str,"\n")

}

##############################################################