**CODE:**

library(tidyverse)

library(naniar)

library(visdat)

library(ggplot2)

library(stringr)

library(dplyr)

library(formattable)

Billionaires <- read.csv("Billionaire.csv", header = TRUE, na.strings = c("NA","N/A",""))

head(Billionaires)

glimpse(Billionaires)

sum(duplicated(Billionaires))

Billionaires %>%

count(Name) %>%

filter (n > 1)

sum(is.na(Billionaires))

miss\_var\_summary(Billionaires)

Billionaires %>%

arrange(NetWorth) %>%

vis\_miss()

Billionaires\_cleaned <- na.omit(Billionaires) %>%

distinct(Name, .keep\_all = TRUE)

sum(is.na(Billionaires\_cleaned))

Billionaires\_cleaned %>%

count(Name) %>%

filter (n > 1)

print(Billionaires$NetWorth)

head(Billionaires\_cleaned)

regexp <- "[[:digit:]]+"

Billionaires\_cleaned$NumNetworth<-str\_extract(Billionaires\_cleaned$NetWorth, regexp)

Billionaires\_cleaned$NumNetworth<-as.numeric(Billionaires\_cleaned$NumNetworth)

Billionaires\_cleaned<-Billionaires\_cleaned[order(Billionaires\_cleaned$NumNetworth, decreasing = TRUE),]

topten<-head(Billionaires\_cleaned,10)

topten

ggplot(topten,aes(x= reorder(Name, desc(NumNetworth))))+geom\_bar(aes(fill= NetWorth))+labs(title = "Top 10 Billionaires in World")+theme(axis.text.x=element\_text(angle=90,hjust=1,vjust=0.5))+labs(x="Names",y="Count")

# Top Five Industry with most Billionaires.

Industry<-as.data.frame(table(Billionaires\_cleaned['Industry']))

colnames(Industry)<-c("Industry","Occurences")

Industry$Fraction <- Industry$Occurences / sum(Industry$Occurences)

Industry

Industry<-Industry[order(Industry$Occurences, decreasing = TRUE),]

Industry

TopFiveIndustry<-head(Industry,5)

TopFiveIndustry$Fraction <- TopFiveIndustry$Occurences / sum(TopFiveIndustry$Occurences)

TopFiveIndustry

TopFiveIndustry$max <- cumsum(TopFiveIndustry$Fraction)

TopFiveIndustry$min <- c(0, head(TopFiveIndustry$max, n=-1))

TopFiveIndustry

TopFiveIndustry$labelPosition <- (TopFiveIndustry$max + TopFiveIndustry$min) / 2

TopFiveIndustry$label <- paste0(TopFiveIndustry$Industry, "\n value: ", formattable(TopFiveIndustry$Fraction\*100,digits = 2, format = "f"))

TopFiveIndustry

ggplot(TopFiveIndustry, aes(ymax=max, ymin=min, xmax=4, xmin=3, fill=Industry)) +

geom\_rect() +

geom\_label( x=3.5, aes(y=labelPosition, label=label), size=6) +

scale\_fill\_brewer(palette=8) +

coord\_polar(theta="y") +

xlim(c(2, 4)) +

theme\_void() +

theme(legend.position = "none")

# Top Five Countries of Billionaires.

Country<-as.data.frame(table(Billionaires\_cleaned['Country']))

colnames(Country)<-c("Country","Occurences")

Country$Fraction <- Country$Occurences / sum(Country$Occurences)

Country

Country<-Country[order(Country$Occurences, decreasing = TRUE),]

Country

TopFiveCountry<-head(Country,5)

TopFiveCountry$Fraction <- TopFiveCountry$Occurences / sum(TopFiveCountry$Occurences)

TopFiveCountry

TopFiveCountry$max <- cumsum(TopFiveCountry$Fraction)

TopFiveCountry$min <- c(0, head(TopFiveCountry$max, n=-1))

TopFiveCountry

TopFiveCountry$labelPosition <- (TopFiveCountry$max + TopFiveCountry$min) / 2

TopFiveCountry$label <- paste0(TopFiveCountry$Country, "\n value: ", formattable(TopFiveCountry$Fraction\*100,digits = 2, format = "f"))

TopFiveCountry

ggplot(TopFiveCountry, aes(ymax=max, ymin=min, xmax=4, xmin=3, fill=Country)) +

geom\_rect() +

geom\_label( x=3.5, aes(y=labelPosition, label=label), size=6) +

scale\_fill\_brewer(palette=8) +

coord\_polar(theta="y",start = 90) +

xlim(c(2, 4)) +

theme\_void() +

theme(legend.position = "none")

#Top Five Domains of Billionaires.

Domains<-as.data.frame(table(Billionaires\_cleaned['Source']))

colnames(Domains)<-c("Source","Occurences")

Domains$Fraction <- Domains$Occurences / sum(Domains$Occurences)

Domains

Domains<-Domains[order(Domains$Occurences, decreasing = TRUE),]

Domains

TopFiveDomains<-head(Domains,5)

TopFiveDomains$Fraction <- TopFiveDomains$Occurences / sum(TopFiveDomains$Occurences)

TopFiveDomains

TopFiveDomains$max <- cumsum(TopFiveDomains$Fraction)

TopFiveDomains$min <- c(0, head(TopFiveDomains$max, n=-1))

TopFiveDomains

TopFiveDomains$labelPosition <- (TopFiveDomains$max + TopFiveDomains$min) / 2

TopFiveDomains$label <- paste0(TopFiveDomains$Source, "\n value: ", formattable(TopFiveDomains$Fraction\*100,digits = 2, format = "f"))

TopFiveDomains

ggplot(TopFiveDomains, aes(ymax=max, ymin=min, xmax=4, xmin=3, fill=Source)) +

geom\_rect() +

geom\_label( x=3.5, aes(y=labelPosition, label=label), size=6) +

scale\_fill\_brewer(palette=8) +

coord\_polar(theta="y",start = 90) +

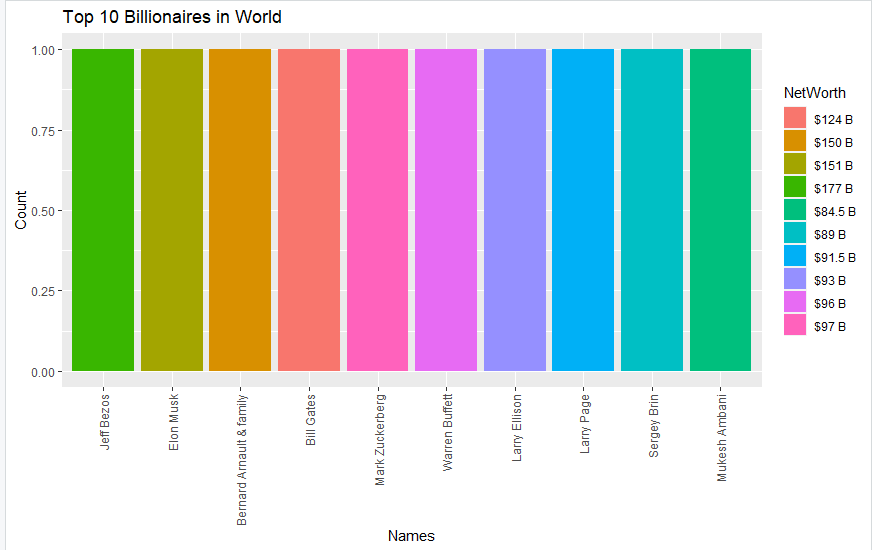
xlim(c(2, 4)) +

theme\_void() +

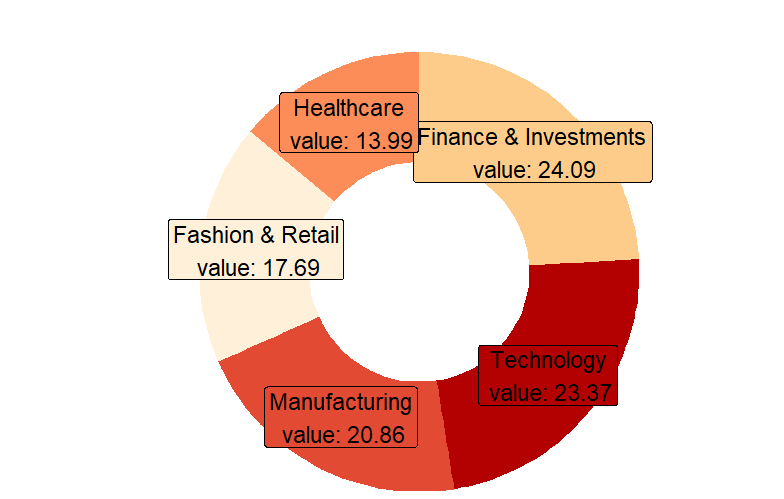
theme(legend.position = "none")

**Screen Shots:**

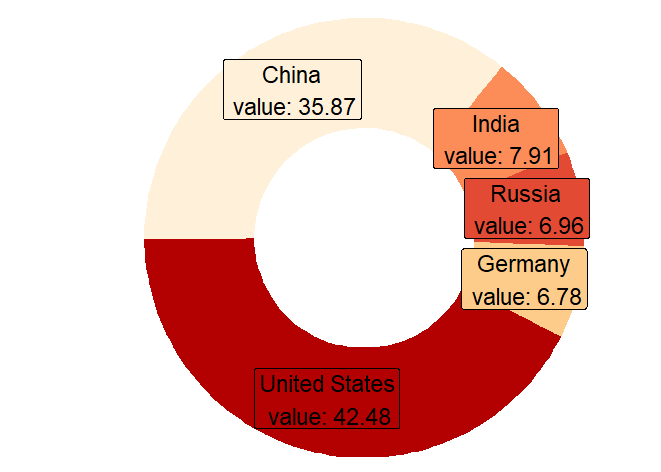




**Top 5 Industries with most Billionaires**



**Top Five countries with most Billionaires**



**Top Five Domains with most Billionaires**

