M Hamza Ihsan

19F0107 | NUCES

Assignment 3

Exploratory Data Analysis

**EDA TASKS:**

library(tidyverse)

library("rvest")

url<-"https://docs.google.com/spreadsheets/d/e/2PACX-1vQA9Atb2iCAbExo6bGyneP0yYIwaHn3xtaPSOATzziQqELAbBy0hp-bWjC2Y2HWbxAbHsZMr7NmF14s0S-pcEQ/pub"

page <- read\_html(url)

tabs <- html\_nodes(page, xpath = "//li")

tab\_df <- data.frame(

name = tabs %>% html\_text,

css = paste0("#", sub("\\D", "", html\_attr(tabs, "id")))

)

data<-rvest::read\_html(url) %>%

rvest::html\_nodes("table") %>%

.[1:9]%>%

rvest::html\_table(fill=TRUE)

matrixx<-data.frame(data[6])

matrixx\_july<-data.frame(data[7])

matrixx[,34]

matrixx\_july[,34]

d<-matrixx[,34]

d\_july<-matrixx\_july[,34]

mat1.data<-as.numeric(c(d\_july))

mat1<-matrix(mat1.data,nrow=152)

mat1<-mat1[-c(1),]

mat2.data<-as.numeric(c(d))

mat2<-matrix(mat2.data,ncol=153)

mat2<-mat2[,-c(1,2)]

mat3<-list()

length(mat1)

length(mat2)

for(var in 1:151){

mat3[[(var+1)]]<-(mat1[var]-mat2[var]/100)

}

dat<-data.frame(mat1,mat2)

ggplot(dat)+

geom\_jitter(aes(x=mat1,y=mat2),alpha=0.2,width=0.1)+

stat\_summary(aes(x=mat1,y=mat2),fun=median,color='blue' )

ggplot(dat)+

geom\_density(aes(x=mat1),position="identity",alpha=0.4)

ggplot(dat)+

geom\_histogram(aes(x=mat1),position="identity",alpha=0.4)

ggplot(dat)+

geom\_density(aes(x=mat2),position="identity",alpha=0.4)

ggplot(dat, aes(x=mat2,y=mat1,position = "identity"))+

geom\_density\_2d() +geom\_point() #Not really helpfull

