Reading pdf from R

**Getting a list of documents in a folder**

file\_vector <- list.files(path = "data")

head(file\_vector)

**Get the list of pdf files in a folder**

grepl(".pdf",file\_vector)

**Extract only pdf files**

pdf\_list <- file\_vector[grepl(".pdf",file\_vector)]

head(pdf\_list)

**Install package pdftools [Note: only for one time]**

install.packages("pdftools")

**Include the package in the R environment:**

library("pdftools")

pdf\_text("data/geo plitics main ppt.pdf")

pdftxt = pdf\_text("data/geo plitics main ppt.pdf")

splitpdftxt = strsplit(pdftxt, split = "\n")

print(splitpdftxt)