**R + Py**

In the word of R vs Python fights, This is a simple (could be called, naive as well) attempt to show how we can combine the power of Python with R and create a new superpower.

Like this one, If you have watched **The Incredibles** before!

**About this Dataset**

This dataset contains a bunch of tweet that came with this tag **#JustDoIt** after **Nike** released the ad campaign with Colin Kaepernick that turned controversial.

Dataset source: <https://www.kaggle.com/eliasdabbas/5000-justdoit-tweets-dataset>

**Superstar – Reticulate**

The superstar who’s making this possible is the R package reticulate by RStudio.

*Let us start with the code!!*

**The R Code**

#loading required R libraries

library(tidyverse)

library(ggthemes)

library(knitr)

tweets <- read\_csv("<https://raw.githubusercontent.com/amrrs/python_plus_r_brug/master/justdoit_tweets_2018_09_07_2.csv>")

text <- tweets$tweet\_full\_text

set.seed(123)

text\_10 <- text[sample(1:nrow(tweets),100)]

**The Python Code**

import spacy

import pandas as pd

nlp = spacy.load('en\_core\_web\_sm')

doc = nlp(str(r.text\_10))

pos\_df = pd.DataFrame(columns = ["text","pos","lemma"])

for token in doc:

df1 = pd.DataFrame({"text" : token.text, "pos" : token.pos\_, "lemma" : token.lemma\_}, index = [0])

#print(token.text, token.pos\_)

#print(df1)

pos\_df = pd.concat([pos\_df,df1])

#print(pos\_df)

**Now, Again The R Code**

#data.frame(token = as.vector(py$tokens)) %>% count(token) %>% arrange(desc(n))

py$pos\_df %>%

count(pos) %>%

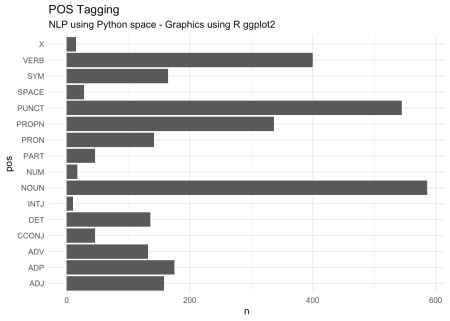
ggplot() + geom\_bar(aes(pos,n), stat = "identity") +

coord\_flip() +

theme\_minimal() +

labs(title = "POS Tagging",

subtitle = "NLP using Python space - Graphics using R ggplot2")



**Now, Again The Python Code**

ent\_df = pd.DataFrame(columns = ["text","label"])

for ent in doc.ents:

df1 = pd.DataFrame({"text" : ent.text, "label" : ent.label\_}, index = [0])

#print(token.text, token.pos\_)

#print(df1)

ent\_df = pd.concat([ent\_df,df1])

**One Final Time, The R Code**

py$ent\_df %>%

count(label) %>%

ggplot() + geom\_bar(aes(label,n), stat = "identity") +

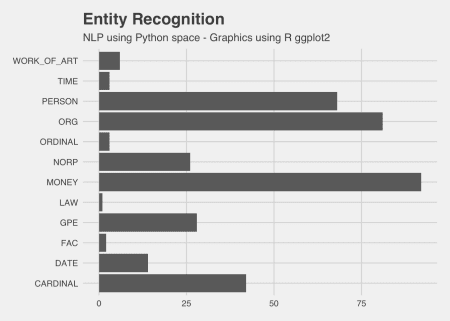
coord\_flip() +

#theme\_solarized() +

theme\_fivethirtyeight() +

labs(title = "Entity Recognition",

subtitle = "NLP using Python space - Graphics using R ggplot2")



**Summary**

Thus, In this post we learnt how to combine the best of R and Python – in this case – R for Data Analysis and Data Visualization – Python with Spacy.