

# Exam code R PSB

Jeremie Sayag

1/27/2021

```
install.packages("reticulate", repos = "http://cran.us.r-project.org")
```

```
##  
## The downloaded binary packages are in  
## /var/folders/dk/_s4_y6ps10lgt33j5nzwz42r0000gn/T//Rtmpsyiuy0/downloaded_packages
```

Criteres : Qualite de la presentation, pedagogie, decouverte, utilisation, simplicité

## INTRODUCTION A GIT

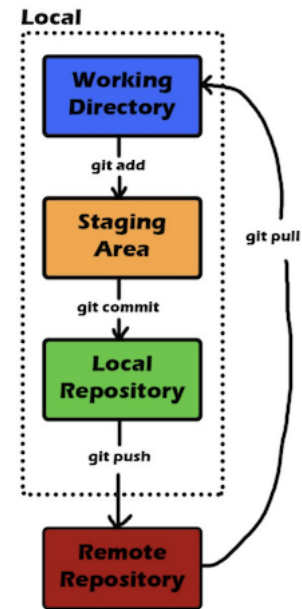
lien github : [https://github.com/soukainaElGhaldy/PSBX\\_tutoGit-GitHub/tree/master/Git\\_Github\\_Documentation](https://github.com/soukainaElGhaldy/PSBX_tutoGit-GitHub/tree/master/Git_Github_Documentation)  
Nom des auteurs : Adrien JUPITER & Soukaina EL GHALDY & Siva CHANEMOUGAM

### What's Inside

Ce travail explique comment utiliser github et ses differentes commandes via un ordinateur. Github est une plateforme web qui permet de stocker du code et facilite le travail de groupe. Git est un scv local c'est à dire un systeme de controle de version, qui permet de travailler avec github directement via notre ordinateur personnel en important et exportant des lignes de code depuis le terminal de notre ordinateur très rapidement.

### Focus

Comment faire pour envoyer notre code sur github et permettre une collaboration en equipe ? Ce tuto nous l'explique : Il tout d'abord creer un repertoire git depuis un dossier local (depot git) avec la commande "git init". Puis avec la commande "git add" nous preparons les fichiers que nous voulons envoyer sur github, ils sont en -staging area-. Ensuite, avec la commande "git commit -m" nous preparons la version du fichier à envoyer avec la description de cette version. A ce moment, tout est pret au niveau local (local repository) et nous avons la derniere version du fichier à envoyer. Enfin, nous envoyons le fichier sur github sur notre



remote repository. Le schema ci-dessous explique parfaitement cela.

## V Criterias

- Qualité de la presentation : La presentation est dans son ensemble claire, utilisation d'images et schemas pour nous aider à comprendre
- Pedagogie : Les auteurs ont reussi à nous faire comprendre l'utilisation de git et Github -Decouverte : Git et Github sont tres connus et incontournable dans le monde du digital. -Utilisation : Ce tuto est tres utile puisque des millions de personnes dans le monde utilisent Git et Github quotidiennement. -Simplicité : Ce tuto est a la porté de tout le monde du simple padawan au meilleur des jedi

## The End

J'ai beaucoup apprecié le travail des auteurs qui expliquent de maniere tres didactique l'incontournable plateforme Github et comment utilise Git.

---

## INTRODUCTION A LINUX

lien github : <https://github.com/gfontainepsb/Cours-R> Nom des auteurs : FONTAINE Grégoire

## What's Inside

Apres une introduction sur le systeme Linux, l'auteur nous explique pas à pas comment installer Ubuntu (un systeme Linux) sur notre ordinateur et ce quelque soit notre systeme d'exploitation (mac os ou microsoft). L'auteur nous explique comment faire cette installation via une machine virtuelle (virtual box) sur notre ordinateur.

## Focus

Une machine virtuelle ou virtual box est en quelque sorte un ordinateur (virtuel) dans un autre ordinateur. Nous pouvons donc faire tourner un système d'exploitation Linux sur un même ordinateur avec Windows ou un Mac. Pour cela, il faut : 1) Télécharger la virtual box, <https://www.virtualbox.org/> 2) Installer la virtual box sur notre ordinateur 3) Installer le système Linux qui nous convient (Ubuntu ici), <https://ubuntu.com/> 4) Paramétrer Virtual box avec Ubuntu en suivant les instructions ici : <https://fr.wikihow.com/installer-Ubuntu-sur-VirtualBox>

## V Criterias

- Qualité de la présentation : Présentation claire et succincte
- Pédagogie : En un temps court l'auteur nous guide à travers l'installation d'Ubuntu de façon très pédagogique. - Découverte : N'ayant jamais utilisé une virtual box et ubuntu cela est une découverte pour moi - Utilisation : Linux est un système très utilisé dans le monde et notamment dans la sécurité informatique, ce guide sur l'installation de Linux sur son ordinateur est utile à un grand nombre de personnes. - Simplicité : Ce guide pas à pas nous tient littéralement par la main pour installer Linux sur notre machine.

## The End

J'ai aimé le travail de l'auteur, qui ne tourne pas autour du pot et rentre dans le concret directement avec ce guide étape par étape pour installer Linux sur notre ordinateur.

---

## EXTRAIRE LE CONTENU D'UN PDF AVEC R

lien github : <https://github.com/chaymae-data/PSBX> Nom des auteurs : GASMI Chaymae

## What's Inside

L'auteur nous explique comment extraire le contenu d'un fichier pdf en R au moyen de 2 packages R différents : Package pdftools et package tm. Puis, une fois le contenu extrait, comment nettoyer ce dernier afin de pouvoir lancer des analyses sémantiques dessus.

## Focus

Utilisons le package "pdftools" afin d'extraire le contenu d'un fichier pdf en R. Pour cela nous devons : 1) Installer ce package via le bouton install sur la fenêtre en bas à droite sur RStudio. 2) Charger le package dans notre code en R : `library(pdftools)`

```
#install.packages(pdftools)
library(pdftools)
```

```
## Using poppler version 0.73.0
```

3) On importe depuis le site voulu le pdf

```
#install.packages(pdftools)
library(pdftools)
download.file("https://www.btbores.org/Downloads/I%20Have%20a%20Dream%20by%20Martin%20Luther%20King%20Jr.pdf", mode = "wb")
```

4) On le charge dans R au moyen de “pdf\_text” 5) On utilise “strsplit” pour séparer les lignes les unes des autres 6) Enfin, on utilise la fonction “cat” qui affiche de façon simple les résultats sous forme de texte dans la console et permet l’export des résultats dans un objet

```
#install.packages(pdftools)
#library(pdftools)
#download.file("https://www.btbores.org/Downloads/I%20Have%20a%20Dream%20by%20Martin%20Luther%20King%20Jr.pdf", mode = "wb")
#text <- pdf_text("I%20Have%20a%20Dream%20by%20Martin%20Luther%20King%20Jr.pdf")
#text1 <- strsplit(text, "\n")
#cat(text[1])
```

## V Critierias

- Qualité de la presentation : Presentation tres complete
- Pedagogie : L’auteur nous explique à l’aide d’insertion de code R dans sa presentation comment utiliser ces differents packages. -Decouverte : C’est une complete decouverte pour moi -Utilisation : Le pdf est une extension tres utilise pour les fichiers texte. Il est donc tres utile de savoir comment extraire et traiter les donnees en code R. -Simplicité : L’auteur de façon tres complete nous explique l’utilisation de ces packages, cependant il est possible que les veritables novices en R puissent comprendre l’integralite des explications proposees.

## The End

Il est tres utile de savoir manipuler des fichiers pdf qui sont tres repandus. J’apprecie donc grandement le travail de l’auteur qui nous explique cela etape par etape pour apres pouvoir utiliser ces donnees.

## R vs Python

lien github : [https://github.com/Cldren/REN\\_PSBx/tree/main/R\\_Docs](https://github.com/Cldren/REN_PSBx/tree/main/R_Docs) Nom des auteurs : Claude REN

## What’s Inside

Une comparaison des operations mathematiques entre le langage R et le langage Python. Une comparaison d’operations mathematiques basiques, la manipulation de vecteurs sur en R et Python. L’équivalence de la fonction str(u) en Python, ou plutot le manque d’équivalence de cette fonction venu de R qui nous oblige a cree une fonction specifique en Python.

## Focus

Faisons un focus sur la fonction str() en R et sa difference en Python.

```
str(u) num [1:6] 5 5 5 32 62 49 import pandas as pd du = pd.DataFrame(u) print(str(du.info()) + "n"+str(u))
```

Voici le resultat de ce code python :

```
<class 'pandas.core.frame.DataFrame'> RangeIndex: 6 entries, 0 to 5 Data columns (total 1 columns):  
Column Non-Null Count Dtype --- --- --- 0 0 6 non-null int64 dtypes: int64(1) memory usage:  
176.0 bytes None [ 5 5 5 32 62 49]
```

C'est donc une fonction totalement differente de ce qu'on trouve en R. En Python cette fonction permet un type non string en string.

## V Criterias

- Qualité de la presentation : Presentation structure et comparatif entre le code Python et R
- Pedagogie : Comparaison verticale tres facile a comprendre ligne par ligne. -Decouverte : La difference d'utilisation de la fonction str() est une totale decouverte -Utilisation : Excellent fichier a garder pour avoir une vue rapide de different code en python et R. -Simplicité : Assez simple mis a part certain concept mathematiques qui peuvent etre complique a comprendre.

## The End

Travail tres didactique et comparaison tres utile a avoir sous la main lorsque l'on travail dans le domaine des Data Science.

---

## MANIPULATION DE STRING AVEC R

lien github : <https://github.com/LeoBsn/PSB-PROJECT> Nom des auteurs : Léonard BOISSON

## What's Inside

Une presentation, explication, analyse et tuto du package "stringr" développé par Hadley Wickham. Ce package permet de simplifier les fonctionnalites de R appliquees au chaine de caracteres (string). Cela permet de simplifier l'utilisation de string dans le code R, de faire du text mining grace a ces donnees et de récupérer des infos emprisonnées dans des chaînes de caractères en provenance d'un fichier excel.

## Focus

Comment utiliser cette fonction "stringr" dans RStudio ? Il faut tout d'abord installer le package grace au bouton install dans la fenetre en bas a droite de RStudio. Puis, charger cette fonction dans R en utilisant la commande suivante :

```
library('stringr')
```

Importons ensuite le fichier a analyser grace a la fonction "read\_excel"

```
library(readxl)
```

Puis, de la meme maniere chargeons notre fichier avec cette meme fonction :

```
#d <- read_excel("C:/Users/leona/netflix_titles.xlsx", col_types = c("text",
#"text", "text", "text", "text",
#"text", "text", "text", "text", "text",
#"text"))
```

Notre fichier est pret a etre manipuler ! Toutes les manipulations possibles sont resumés dans ce tres bon schema apporte par l'auteur.

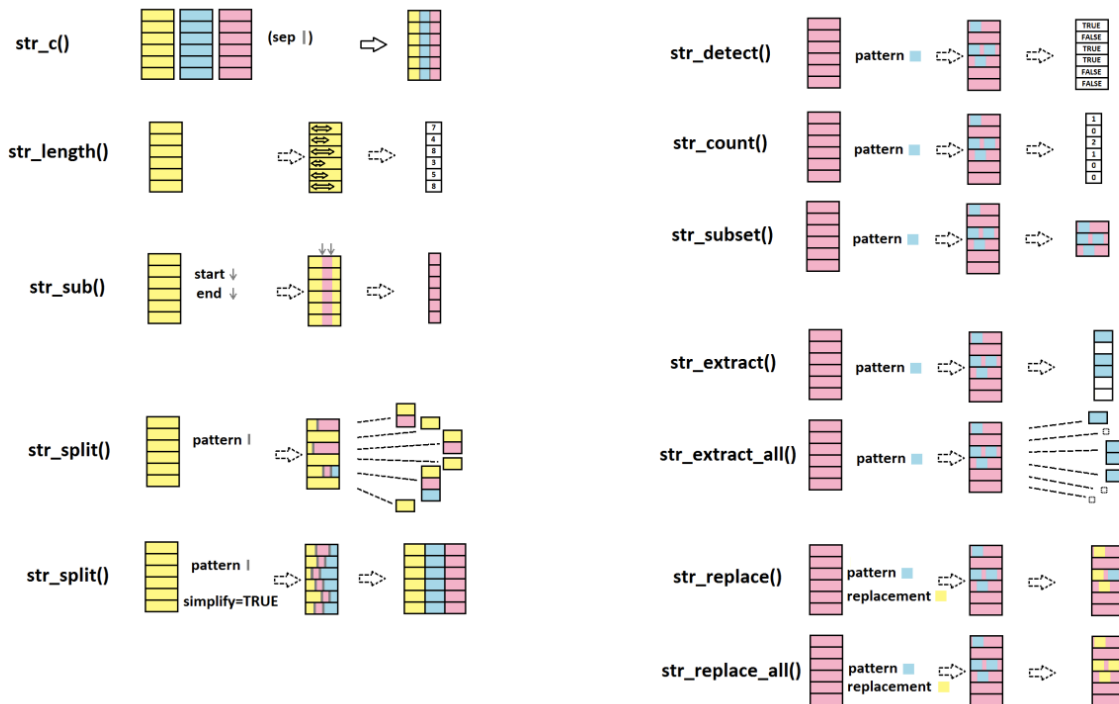


Figure 1: tuto git

## V Criterias

- Qualité de la presentation : Claire, Efficace dans son contenu
- Pedagogie : Utilisation de schema pour expliquer le contenu ainsi que de resultat de la console et bibliographie tres complete. -Decouverte : J'avais deja traite des donnees avec en excel avec python jamais en R -Utilisation : Excel restant un must pour le traitement de donnees, il est necessaire de savoir comment les traiter en R, et savoir comment traiter les donnees caracteres. -Simplicité : Le plus dure etant de comprendre comment R lit le fichier excel, les differents sens et apres cela devient facile.

## The End

R etant le language de predilection de la data science, et excel son pendant en fichier la liaison entre les 2 parait necessaire, et ce tuto prend tout son sens. Merci a l'auteur.

## MON PACKAGE : DATAEXPLORER

lien github : <https://github.com/Jeremiesayag/PSBX-R> Nom des auteurs : Jeremie SAYAG

### What's Inside

Une presentation, explication, analyse et tuto du package “dataexplorer” . Ce package permet d’explorer les données issues d’un dataset. Il nous permet d’analyser les distributions des variables ou encore de visualiser de facon méthodique et minutieuse notre de jeu de données.

### Focus

Voyons un exemple de ce package. Chargeons le package et utilisons le sur un dataset connu de R que l’on appelle “iris” qui donne les mesures en centimètres des variables longueur et largeur des sépales et longueur et largeur des pétales, respectivement, pour 50 fleurs de chacune des 3 espèces d’iris. Et regardons plus en détail ce dataset le grace a la fonction head.

```
library(datasets)
# Vue rapide dataset grace à la fonction head()
head(iris)
```

##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
## 1	5.1	3.5	1.4	0.2	setosa
## 2	4.9	3.0	1.4	0.2	setosa
## 3	4.7	3.2	1.3	0.2	setosa
## 4	4.6	3.1	1.5	0.2	setosa
## 5	5.0	3.6	1.4	0.2	setosa
## 6	5.4	3.9	1.7	0.4	setosa

Nous pouvons également visualiser notre dataset grace a la fonction plot.

### V Criterias

- Qualité de la presentation : Succinte
- Pedagogie : bonne visualisation grace a la fonction plot qui permet de mieux comprendre -Decouverte : bonne decouverte pour moi -Utilisation : Traiter les donnees est l’objectif de R, en ce sens le package DataExplorer permet d’atteindre cet objectif. -Simplicité : Le plus dure est de comprendre le fonctionnement du jeu de donnee, ensuite l’utilisation du package est plutot intuitif.

### The End

J’ai essaye de faciliter le plus possible l’utilisation du package “DataExplorer” que je trouve tres utile pour la visualisation et la comprehension du de dataset en R.

##The Real End

La plupart de mes camarades ont comme moi choisi de traiter un package ayant trait aux jeux de donnees. J’ai aime l’approche tres didactique etape par etape utilisee par certains qui permettent de suivre scrupuleusement le tuto et d’arriver a le faire par soi-meme bien que n’ayant que peu de connaissance concernant le package ou meme aucune. Il est possible de plus detailler le package que j’ai explique “DataExplorer” et d’utiliser d’autres dataset afin de comprendre plus en profondeur l’utilisation de Data en R.

Les travaux donnés par notre professeur Mr Laude et le travail de recherche effectués pour comprendre, analyser et commenter les travaux de mes camarades ainsi que les miens m'ont permis d'étendre ma connaissance sur le langage R, langage de prédilection en Data Science.