Référence Github/Travaux/Packages Promo MSc Data Management

Soukaina EL GHALDY & Antoine SERREAU

---

1. Avant propos :

Afin d'accompagner l'ensemble de la promotion du MSc Data Management à Paris School of Business dans le suivi du cours du Professeur intervenant **M. Henri LAUDE** dans les deux matières **"Mathématiques pour le Big Data"** et **"Programmation en R"**. Les délégués du MSc Data management ont créé ce fichier afin de vous aider à vous repairer par rapport aux attentes du cours.

Avant de commencer, n'oubliez pas de parcourir attentivement toutes les ressources et visualiser la vidéo de 20 minutes sur le Fact-checking qui sont sur le GitHub du Professeur LAUDE.

Vous trouverez dans ce document plusieurs tableaux qui vous guideront et vous aiderons à réaliser vos recherches dans cette « base de connaissances », notamment lors des évaluations.

2. Les consignes à suivre :

Dans un premier temps, vous allez devoir créer obligatoirement un compte GitHub. Vous communiquerez ensuite votre identifiant Github aux délégués. Ensuite, sur votre compte GitHub, vous allez « **Fork-er** » à partir du GitHub du professeur le dépôt « **PSB1** ».

Bravo ! Vous venez d'effectuer votre première manipulation GitHub.

Dans un deuxième temps, vous créerez un dépôt (ou « **repository** » en anglais) que vous appellerez « **PSB-X** ». Ce dépôt sera l'environnement où vous rendrez par la suite l'ensemble de vos travaux.

Ci-dessous, la liste des étudiants avec leurs adressses mails et le nom du Github associé.

Table 1: Liste Etudiants/Mail/Github

Nom	Prénom	Mail	Github_ID
ABBES	*Ahmed*	a_abbes@stu-psbedu.paris	**Ahmed-Abbes5**
ALLAKER	*Maxime*	m_allakerehormo@stu-psbedu.paris	**mallaker**
ALLIX	*Nicolas*	n_allix2@stu-psbedu.paris	**Nicolas-all**
ARSIC	*Marko*	m_arsic@stu-psbedu.paris	**ArsicMPSB**
AUFRERE	*Thuy*	t_aufrere@stu-psbedu.paris	**T-AUF**
BENSALEM	*Akram*	a_bensalem1@stu-psbedu.paris	**AkramBensalemPSB**
BEN YOUSSEF	*Salah*	s_benyoussef1@stu-psbedu.paris	**Salah1920**
BILLAUD	*Lucas*	l_billaud@stu-psbedu.paris	**lucasblld**
BOISSON	*Léonard*	l_boisson@stu-psbedu.paris	**LeoBsn**
BRAHAM	*Ahmed*	a_braham@stu-psbedu.paris	**AhBraham**
BRETONNIERE	*Corentin*	c_bretonniere@stu-psbedu.paris	**CBRETONNIERE**
CHANEMOUGAM	*Siva*	s_chanemougam@stu-psbedu.paris	**Siva-chane**
COMLAN	*Florine*	f_comlan@stu-psbedu.paris	**fcom-stack**

Table 1: Liste Etudiants/Mail/Github *(continued)*

Nom	Prénom	Mail	Github_ID
DAIF	*Hakim*	h_daif@stu-psbedu.paris	**hakim-daif**
DANYACH	*Marion*	m_danyach@stu-psbedu.paris	**MarionD436**
DERROUICHE	*Imen*	i_derrouiche@stu-psbedu.paris	**Imenderrouiche**
EL GHALDY	*Soukaina*	s_elghaldy@stu-psbedu.paris	**soukainaElGhaldy**
FONTAINE	*Gregoire*	g_fontaine3@stu-psbedu.paris	**gfontainepsb**
FORASACCO	*Arnaud*	a_forasacco@stu-psbedu.paris	**ArnaudFrsc**
GASMI	*Chaymae*	c_gasmi@stu-psbedu.paris	**chaymae-data**
GUIGON	*Benjamin*	b_guigon@stu-psbedu.paris	**benjaminguigon**
HOUNSINOU	*Jordy*	j_hounsinou@stu-psbedu.paris	**Jordyhsn**
HOUNTONDJI	*Ramya*	r_hountondji1@stu-psbedu.paris	**RamyaHTDJ**
JUPITER	*Adrien*	a_jupiter@stu-psbedu.paris	**akjupiter**
KANLANFEYI	*Souglman Kabirou*	s_kanlanfeyi@stu-psbedu.paris	**kabirou7**
LAMTI	*Olfa*	o_lamti@stu-psbedu.paris	**OlfaLmt**
LIU	*Jiayue*	j_liu27@stu-psbedu.paris	**liu-jiayue**
LUTZ	*Rindra*	r_lutz@stu-psbedu.paris	**rindra-lutz**
MASSE	*Thomas*	t_masse@stu-psbedu.paris	**Thomas-MAS**
MAZZUCATO	*Claire*	c_mazzucato@stu-psbedu.paris	**clairemazzucato**
PALAY	*Gaspard*	g_palay@stu-psbedu.paris	**GaspardPalay**
REN	*Claude*	c_ren4@stu-psbedu.paris	**Cldren**
RIDADARAJAT	*Zakaria*	z_ridadarajat@stu-psbedu.paris	**zakariaridarajat**
ROBACHE	*William*	w_robache@stu-psbedu.paris	**WilliamRbc**
SAYAG	*Jeremie*	j_sayag1@stu-psbedu.paris	**Jeremiesayag**
SERREAU	*Antoine*	a_serreau@stu-psbedu.paris	**aserreau**
YANKO	*Arnaud Bruel*	a_yankokouatchou@stu-psbedu.paris	**ARNAUDBRUEL-YANKO**
ZOUMANIGUI	*Nina*	n_zoumanigui@stu-psbedu.paris	**Nina809**

Comme tout vos travaux seront sur vos Github, ceux-ci seront visibles par tous les étudiants et donc l'ensemble de ces travaux auxquels s'ajoutent les fichiers que vous venez de « **Fork-er** » constitueront la "base de connaissances" de votre promotion pour ces deux matières.

Attention : Vos travaux doivent obligatoirement être livrés sous forme de .Rmd (Rmarkdown) et PDF (ou beamer) dans l'ensemble des GitHub des membres des groupes.

Dans le cadre de votre formation, et en dehors des partiels, vous aurez 3 travaux à rendre.

a. Un travail supplémentaire :

C'est un travail écrit à rendre sur les GitHub de tous les membres du groupe (si vous êtes en groupe) pour le **16/11/2020** sur un des sujets qui vous a été fournit par le professeur ou que vous avez choisit parmi la liste des travaux reçue par mail le 19 octobre 2020.

Quelques travaux supplémentaires feront l'objet d'une présentation rapide à la classe le **16/11/2020** (non notée), afin que chacun puisse les réutiliser en allant les consulter dans le GitHub de ses camarades. Vous devez alors vous tenir prêt à exposer vos travaux livrés sous forme de .Rmd et PDF (ou beamer) aux autres étudiants. Ils devront permettre à chacun de se faire une idée du sujet traité et de faciliter la lecture votre document, puis de vous poser d'éventuelles questions complémentaires.

Ensuite pour le **20/11/2020**, il faudra faire une vidéo sur le travail supplémentaire. L'ensemble des groupes passeront à l'oral cette fois-ci et l'oral sera noté. Il faudra aussi présenter une slide qui va permettre aux étudiants de comprendre les clés de votre travail, les sources et comment est organisé vos dépôts sur GitHub.

Ci-dessous, la liste des travaux supplémentaires par groupe de travail.

Table 2: Liste des travaux supplémentaire par groupe

Nom	Prénom	Travaux_Supp	Groupe_TS
ABBES	*Ahmed*	Mise en forme 1ère session R + commentaires	1
ALLAKER	*Maxime*	Pandas et dataframe	2
BILLAUD	*Lucas*	Pandas et data frame	2
ALLIX	*Nicolas*	Rattle	3
ROBACHE	*William*	Rattle	3
ARSIC	*Marko*	Cross-validation	4
LUTZ	*Rindra*	Cross-Validation	4
AUFRERE	*Thuy*	Manipulation des facteurs	5
MAZZUCATO	*Claire*	Manipulation de facteurs	5
BRAHAM	*Ahmed*	Compromis biais variance	6
BRETONNIERE	*Corentin*	Dérivation R	7
CHANEMOUGAM	*Siva*	Référent Github : livrer un tuto	8
JUPITER	*Adrien*	Référent Github : livrer un tuto	8
COMLAN	*Florine*	Hadoop	9
HOUNTONDJI	*Ramya*	Hadoop	9
DAIF	*Hakim*	p-value	10
RIDADARAJAT	*Zakaria*	p-value	10
DANYACH	*Marion*	Exposé Data Management at scale	11
DERROUICHE	*Imen*	Data Vizualisation Interactive	12
LAMTI	*Olfa*	Data Vizualisation Interactive	12
EL GHALDY	*Soukaina*	Rapport tâches (Rmarkdown/pdf) + tuto Github	13
SERREAU	*Antoine*	Rapport tâches (Rmarkdown/pdf)	13
FONTAINE	*Gregoire*	Référent Linux :Linux pour les nuls et vi/vim	14
FORASACCO	*Arnaud*	Package Forecast et séries temporelles	15
ZOUMANIGUI	*Nina*	Package Forecast et séries temporelles	15
GASMI	*Chaymae*	Extraire le contenu d'un pdf avec R	16
GUIGON	*Benjamin*	Shiny	17
HOUNSINOU	*Jordy*	MNIST et Fashion	18
KANLANFEYI	*Souglman Kabirou*	MNIST et Fashion	18
LIU	*Jiayue*	Tuto LateX	19
MASSE	*Thomas*	Create a map with R	20
PALAY	*Gaspard*	Travail sur les dates avec le package lubridate	21
REN	*Claude*	Python Vs R : syntaxes comparées	22
YANKO	*Arnaud Bruel*	Dossier Time series /ARIMA GARCH	23
SAYAG	*Jeremie*	Variables	24
BEN YOUSSEF	*Salah*	Tests statistiques en R	25
BENSALEM	*Akram*	Sparkr	26
BOISSON	*Léonard*	Stringr	27

b. Un dossier sur les packages R :

Un travail écrit en groupe est à rendre sur les GitHub de tous les membres du groupe pour au plus tard **mi-décembre** (nous nous concerterons pour fixer une date exacte avec le professeur). Ce travail portera sur 2 à 3 packages de R (en dehors des packages déjà évoqués qui correspondent aux travaux supplémentaires).

Attention, il est préférable de ne pas choisir des packages déjà choisis par vos camarades et votre choix doit être soumis avant **16/11/2020**.

Table 3: Liste des travaux sur les Packages R par groupe

Nom	Prénom	Packages_Travaux_R	Groupe_Travaux_R
BRAHAM	*Ahmed*	dplyr / Leaflet	1
EL GHALDY	*Soukaina*	dplyr / Leaflet	1
LIU	*Jiayue*	dplyr / Leaflet	1
BRETONNIERE	*Corentin*	bdpar / dabr / dm	2
SERREAU	*Antoine*	bdpar / dabr / dm	2
ABBES	*Ahmed*	RPART / GGPLOT / gdata	3
BENSALEM	*Akram*	RPART / GGPLOT / PivotalR	3
BEN YOUSSEF	*Salah*	RPART / GGPLOT / PivotalR	3
AUFRERE	*Thuy*	evir / evd / R.minner / graphics	4
YANKO	*Arnaud Bruel*	evir / evd / R.minner / graphics	4
ZOUMANIGUI	*Nina*	evir / evd / R.minner / graphics	4
JUPITER	*Adrien*	Data.table / Parallel / network3D	5
MAZZUCATO	*Claire*	Data.table / Parallel / network3D	5
DAIF	*Hakim*	INFER / Kscorrect / Rstatix	6
GASMI	*Chaymae*	INFER / Kscorrect / Rstatix	6
RIDADARAJAT	*Zakaria*	INFER / Kscorrect / Rstatix	6
DANYACH	*Marion*	Janitor / Plumber	7
DERROUICHE	*Imen*	Janitor / Plumber	7
LAMTI	*Olfa*	Janitor / Plumber	7
LUTZ	*Rindra*	Ggplot2 / Shiny / FactoMineR	8
ROBACHE	*William*	Ggplot2 / Shiny / FactoMineR	8
ALLIX	*Nicolas*	Ggplot2 , Flexdashboard, Tenserflow	9
ARSIC	*Marko*	Ggplot2 , Flexdashboard, Tenserflow	9
ALLAKER	*Maxime*	Ggplot2 / Rpart/ Esquisse	10
CHANEMOUGAM	*Siva*	Ggplot2 / Rpart / Esquisse	10
COMLAN	*Florine*	plymr / rmr2 / sparklyr	11
HOUNTONDJI	*Ramya*	plymr / rmr2 / sparklyr	11
REN	*Claude*	ggplot2 / statsbombR	12
GUIGON	*Benjamin*	Xgboost / e1071 / RandomForest	13
MASSE	*Thomas*	Xgboost / e1071 / RandomForest	13
PALAY	*Gaspard*	Xgboost / e1071 / RandomForest	13
BOISSON	*Léonard*	dply / tidyr / ggplot2	14
FONTAINE	*Gregoire*	dply / tidyr / ggplot2	14
FORASACCO	*Arnaud*	dply / tidyr / ggplot2	14
HOUNSINOU	*Jordy*	Sp / Sf	14
KANLANFEYI	*Souglman Kabirou*	Sp / Sf	14
BILLAUD	*Lucas*	dplyr / rcharts / prophet	15
SAYAG	*Jeremie*	dplyr / rcharts / prophet	15

c. Un dossier de mathématiques pour le Big Data :

Ce travail consiste à travailler en groupe sur les aspects mathématiques de 3 articles de recherche que vous choisirez (ce que de nombreux chercheurs nomment des “papiers” de recherche).

Ici, on ne vous demande pas d’étudier des thèses car elles sont beaucoup plus volumineuses mais bien des **« papier »** de recherche. (sauf biensur si une thèse vous semble très intéressante).

Rappel : Seuls les dossiers sont notés, les travaux supplémentaires sont pourtant à effectuer et amélioreront la note des dossiers si ceux-ci ne sont pas bons.

4. Les Annexes :

Table 4: Tableau récapitulatif des actions à réaliser par l'étudiant

Actions	Commentaires
Créer un compte Github	NA
Créer un répertoire PSB X	Environnement dans le quel vous allez livrer vos codes
Envoyez ce répertoire aux délégués	NA
Faire des groupes de 2-3 (maximum)	Chaque groupe aura des travaux différents
Choisir 2-3 packages	NA
Envoyer ces 3 packages au professeur afin qu'il valide	Expliquer pourquoi vous voulez prendre ces 3 packages
Livrer son travail supplémentaire sur github	Pour la prochaine séance R : le 16 Novembre 2020
Réaliser une vidéo "tuto" de votre travail supplémentaire	Pour le 20 novembre 2020
Rendre le dossier sur les packages avant mi-décembre	Modificiation de la date avec le professeur

5. Sources :

Comment lire un fichier Excel ?

https://readxl.tidyverse.org/reference/read_excel.html

Comment faire un rapport automatisé ?

http://larmarange.github.io/analyse-R/rmarkdown-les-rapports-automatises.html