Examen_R_Maths

Olfa LAMTI

16/01/2021

Contents

Introduction

Vous trouverez dans ce document le résumer et l'évaluation de 12 travaux comportant du code R ou comportant des mathématiques. Les 5 critères que je vais évaluer sont :

- Le niveau de compréhension du sujet
- L'explication des notions
- L'illustration à l'aide d'exemple
- L'utilité du package
- La mise en page

Note : Pour avoir accès aux liens des travaux vous devez cliquer sur chacun des titres des travaux

Travaux comportant du code R

1 - GGPLOT - ALLIX Nicolas et ARSIC Marko

Synthèse du travail

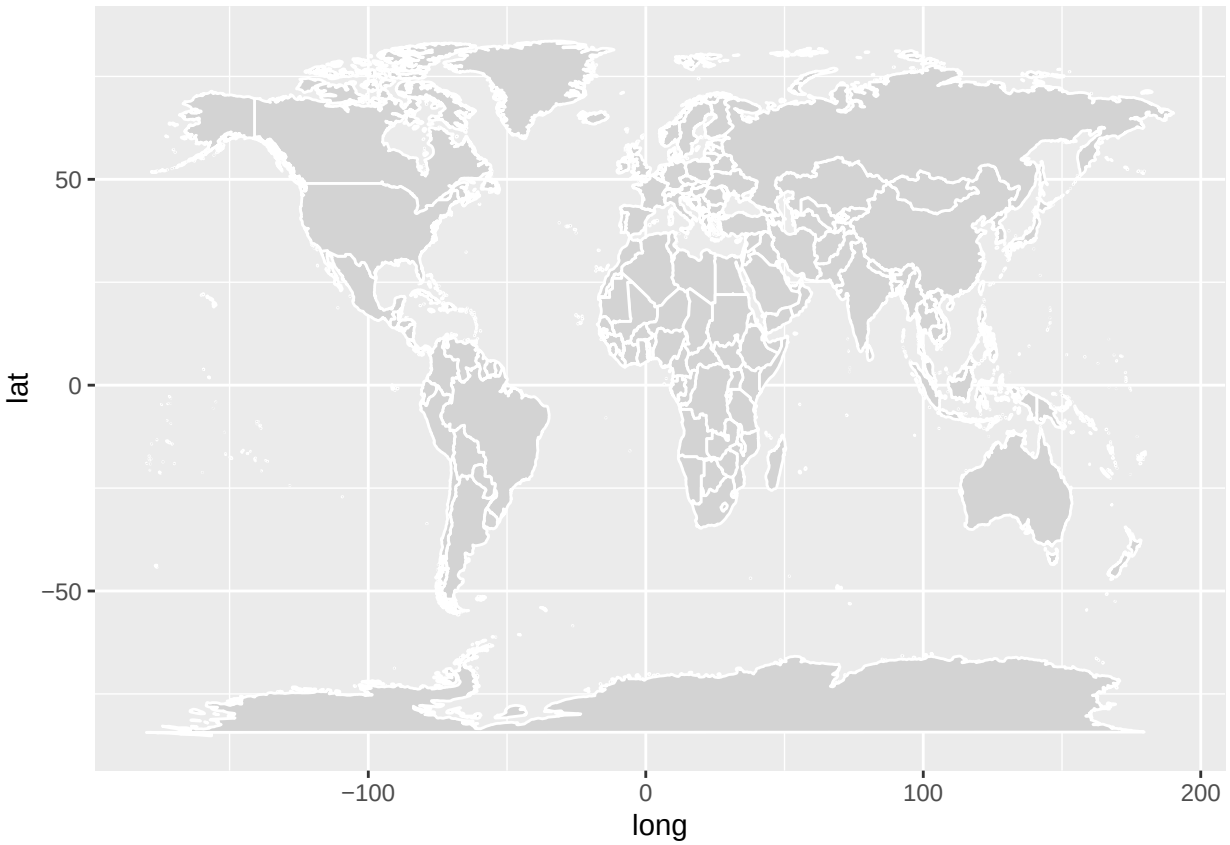
Dans cette partie nous allons analyser le travail de ALLIX Nicolas et ARSIC Marko sur le package GGPLOT2. Ggplot2 est un package de visualisation utilisé aussi bien en R qu'en python. On l'utilise pour faire des graphiques à partir d'une base de données, mais on peut s'en servir pour bien plus que cela.

Extrait commenté des parties

Je vais commenter dans l'utilisation de la carte via gplot2

```
install.packages("maps")
library(maps)
install.packages("mapdata")
library('mapdata')
install.packages('dplyr')
library('dplyr')
install.packages("viridis")
library('viridis')
library('ggplot2')

world_map <- map_data("world")
ggplot(world_map, aes(x = long, y = lat, group = group)) +
  geom_polygon(fill="lightgray", colour = "white")
```



La fonction map_data() dans ggplot2 permet de récupérer les données cartographiques (Nécessite le package maps).

“word” est une carte du monde avec comme données la longitute et latitude.

La fonction geom_polygon() dans ggplot2 permet de créer la carte.

La package viridis pemret de pour définir la palette de couleurs de la carte choroplèthe.

Le resultat nous permet d’avoir une carte du monde en blanc et gris.

Evaluation

- Le niveau de compréhension du sujet : le sujet est assez simple et très clair
- L’explication des notions : les notions sont très bien expliquées.
- L’illustration à l’aide d’exemple : il y a beaucoup d’illustration ce qui est indispensable à la compréhension de ggplot sur les cartes.
- L’utilité du package : le package est très utile pour utiliser des cartes.
- La mise en page : la mise en page est très bien.

Conclusion

En conclusion, le sujet présenté est très intéressant et très utilisé dans R pour faire dtoute sorte de graphique et des cartes et cela a été très bien expliqué.

2 - DPLYR - LIU Jiayue et EL GHALDY Soukaina

Synthèse du travail

Dans cette partie nous allons analyser le travail de LIU Jiayue et EL GHALDY Soukaina sur le package DPLYR. Ils nous présentent dans un premier temps comment installer DPLYR puis comment manipuler les données avec beaucoup d'exemple sur des jeux de données.

Extrait commenté des parties

Je vais commenter dans la fonction Arrange()

```
library(dplyr)
library(nycflights13)
## Chargement des trois tables
data(flights)
arrange(flights, dep_delay)
```

```
## # A tibble: 336,776 x 19
##     year month   day dep_time sched_dep_time dep_delay arr_time sched_arr_time
##    <int> <int> <int>    <int>          <int>     <dbl>    <int>          <int>
##  1  2013    12     7     2040           2123       -43       40           2352
##  2  2013     2     3     2022           2055       -33     2240           2338
##  3  2013    11    10     1408           1440       -32     1549           1559
##  4  2013     1    11     1900           1930       -30     2233           2243
##  5  2013     1    29     1703           1730       -27     1947           1957
##  6  2013     8     9      729            755       -26     1002            955
##  7  2013    10    23     1907           1932       -25     2143           2143
##  8  2013     3    30     2030           2055       -25     2213           2250
##  9  2013     3     2     1431           1455       -24     1601           1631
## 10  2013     5     5      934            958       -24     1225           1309
## # ... with 336,766 more rows, and 11 more variables: arr_delay <dbl>,
## #   carrier <chr>, flight <int>, tailnum <chr>, origin <chr>, dest <chr>,
## #   air_time <dbl>, distance <dbl>, hour <dbl>, minute <dbl>, time_hour <dttm>
```

Ici on a voulu trier le tableau flights selon le retard au départ croissant

```
arrange(flights, month, dep_delay)
```

```
## # A tibble: 336,776 x 19
##     year month   day dep_time sched_dep_time dep_delay arr_time sched_arr_time
##    <int> <int> <int>    <int>          <int>     <dbl>    <int>          <int>
##  1  2013     1    11     1900           1930       -30     2233           2243
##  2  2013     1    29     1703           1730       -27     1947           1957
##  3  2013     1    12     1354           1416       -22     1606           1650
##  4  2013     1    21     2137           2159       -22     2232           2316
##  5  2013     1    20      704            725       -21     1025           1035
##  6  2013     1    12     2050           2110       -20     2310           2355
##  7  2013     1    12     2134           2154       -20        4             50
##  8  2013     1    14     2050           2110       -20     2329           2355
##  9  2013     1     4     2140           2159       -19     2241           2316
## 10  2013     1    11     1947           2005       -18     2209           2230
## # ... with 336,766 more rows, and 11 more variables: arr_delay <dbl>,
## #   carrier <chr>, flight <int>, tailnum <chr>, origin <chr>, dest <chr>,
## #   air_time <dbl>, distance <dbl>, hour <dbl>, minute <dbl>, time_hour <dttm>
```

Ici on a voulu trier le tableau flights selon le mois, puis selon le retard au départ (avec plusieurs colonnes)

```
arrange(flights, desc(dep_delay))
```

```
## # A tibble: 336,776 x 19
##     year month   day dep_time sched_dep_time dep_delay arr_time sched_arr_time
##    <int> <int> <int>    <int>          <int>     <dbl>    <int>          <int>
##  1  2013     1     9      641            900      1301     1242           1530
##  2  2013     6    15     1432           1935      1137     1607           2120
##  3  2013     1    10     1121           1635      1126     1239           1810
##  4  2013     9    20     1139           1845      1014     1457           2210
##  5  2013     7    22      845           1600      1005     1044           1815
##  6  2013     4    10     1100           1900       960     1342           2211
##  7  2013     3    17     2321            810       911      135           1020
##  8  2013     6    27      959           1900       899     1236           2226
##  9  2013     7    22     2257            759       898      121           1026
## 10  2013    12     5      756           1700       896     1058           2020
## # ... with 336,766 more rows, and 11 more variables: arr_delay <dbl>,
## #   carrier <chr>, flight <int>, tailnum <chr>, origin <chr>, dest <chr>,
## #   air_time <dbl>, distance <dbl>, hour <dbl>, minute <dbl>, time_hour <dttm>
```

Ici on a voulu trier le tableau flights selon selon une colonne par ordre décroissant, on lui applique la fonction desc()

```
arrange(flights, desc(dep_delay))
```

```
## # A tibble: 336,776 x 19
##     year month   day dep_time sched_dep_time dep_delay arr_time sched_arr_time
##    <int> <int> <int>    <int>          <int>     <dbl>    <int>          <int>
##  1  2013     1     9      641            900      1301     1242           1530
##  2  2013     6    15     1432           1935      1137     1607           2120
##  3  2013     1    10     1121           1635      1126     1239           1810
##  4  2013     9    20     1139           1845      1014     1457           2210
##  5  2013     7    22      845           1600      1005     1044           1815
##  6  2013     4    10     1100           1900       960     1342           2211
##  7  2013     3    17     2321            810       911      135           1020
##  8  2013     6    27      959           1900       899     1236           2226
##  9  2013     7    22     2257            759       898      121           1026
## 10  2013    12     5      756           1700       896     1058           2020
## # ... with 336,766 more rows, and 11 more variables: arr_delay <dbl>,
## #   carrier <chr>, flight <int>, tailnum <chr>, origin <chr>, dest <chr>,
## #   air_time <dbl>, distance <dbl>, hour <dbl>, minute <dbl>, time_hour <dttm>
```

Ici on a voulu sélectionner les trois vols ayant eu le plus de retard avec la fonction slice()

Evaluation

- Le niveau de compréhension du sujet : le sujet est assez simple et très clair
- L'explication des notions : les notions sont très bien expliquées, chaque ligne de code est trsè bien détaillé en commentaire par la suite.
- L'illustration à l'aide d'exemple : il y a beaucoup d'illustration ce qui est indispensable à la compréhension de dplyr pour visualiser les jeux de données
- L'utilité du package : le package est très utile pour utiliser des jeux de données
- La mise en page : la mise en page est très très bien.

Conclusion

En conclusion, le sujet présenté me semble indispensalbe à l'utilisation de R et le fait de mettre beaucoup de documentation enrichi les explications. DPLYR ressemble à la logique SQL.

3 - Xgboost - GUIGON Benjamin, MASSE Thomas et PALAY Gaspard

Synthèse du travail

Dans cette partie nous allons analyser le travail de GUIGON Benjamin, MASSE Thomas et PALAY Gaspard sur le package Xgboost. Ils nous présentent dans un premier temps comment installer XGBOOST et nous montrent beaucoup d'exemple.

Extrait commenté des parties

Danss le travail on nous explique comment optimisée l'algorithme d'arbres de boosting de gradient. Le Boosting de Gradient est un algorithme d'apprentissage supervisé dont le principe et de combiner les résultats d'un ensemble de modèles plus simple et plus faibles afin de fournir une meilleur prédiction.

Dans la doc on prépare des datasets et on récupère pour le jeu d'entraînemet les labels

Ensuite il y a la phase de l'hyper parametrage

Puis nous voyons cle résultat de la prédiction et le pourcentage de précision.

Evaluation

- Le niveau de compréhension du sujet : le sujet est assez compliqué mais très clair
- L'explication des notions : les notions sont très bien expliquées, chaque ligne de code est très bien détaillé en commentaire par la suite.
- L'illustration à l'aide d'exemple : il ya des illustrations mais les warning aurait pu être retiré
- L'utilité du package : le package est très utile mettre en œuvre des méthodes de Gradient boosting
- La mise en page : la mise en page est très très bien.

Conclusion

En conclusion, le sujet présenté me semble indispensable à l'utilisation de R dans de la Machine Learning et les Data sciences et me sera très utile pour la suite. Ils nous expliquent bien toute les phases pour entrainer un programme d'apprentissage supervisé avec l'aide de dataset.

4 - FLEXDASHBOARD - ALLIX Nicolas et ARSIC Marko

Dans cette partie nous allons analyser le travail de ALLIX Nicolas et ARSIC Marko sur le package FLEXDASHBOARD. Ils nous présentent dans un premier temps comment installer FLEXDASHBOARD puis comment utiliser les fonctions basiques.

Extrait commenté des parties

Je vais commenter la parties jauges.

Nous avons vu comment mettre en place des jauges. Une jauge affiche une valeur numérique sur un compteur qui court entre les valeurs minimale et maximale spécifiées.

Dans la fonction gauge(), les paramètres :

min, indique la valeur numérique minimale

max, indique la valeur numérique maximale

sectors, indique les secteurs colorés personnalisés (par exemple “success”, “warning”, “danger”). Par défaut, toutes les valeurs sont colorées en utilisant la couleur du thème “succès”.

warning, indique le vecteur numérique à deux éléments définissant la plage de valeurs à colorier comme “avertissement” (couleur spécifique fournie par thème ou personnalisée colors)

danger , indique le vecteur numérique à deux éléments définissant la plage de valeurs à colorier comme “danger” (couleur spécifique fournie par thème ou personnalisée colors)

Evaluation

- Le niveau de compréhension du sujet : le sujet est assez simple et très clair
- L’explication des notions : les notions sont très bien expliquées, chaque ligne de code est très bien détaillé en commentaire par la suite.
- L’illustration à l’aide d’exemple : beaucoup d’illustration
- L’utilité du package : le package est très utile pour faire des dashboard
- La mise en page : la mise en page est très bien.

Conclusion

En conclusion, le sujet présenté me semble indispensable à l’utilisation de R pour faire des dashboard et le fait de mettre beaucoup de documentation enrichi les explications. FLEXDASHBOARD ressemble beaucoup à PLOTLY, le packge que j’ai travaillé.

5 - INFER - DAIF Hakim, GASMI Chaymae et RIDADARAJAT Zakaria

Dans cette partie nous allons analyser le travail de DAIF Hakim, GASMI Chaymae et RIDADARAJAT Zakaria sur le package INFER. Ils nous présentent dans un premier temps les fonction principales de INFER puis les points forts des packages puis des exemples

Extrait commenté des parties

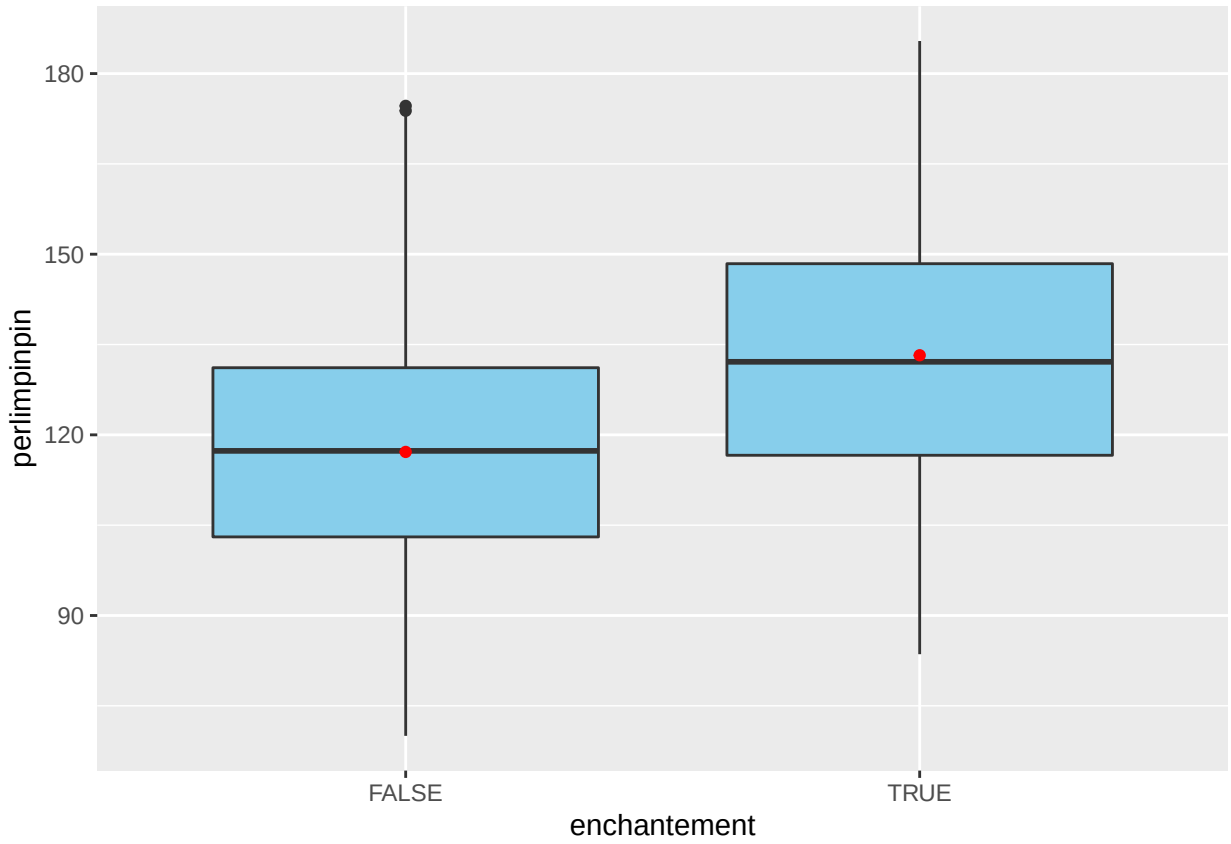
Je vais commenter le bout de code suivant :

```
# connection et mise en place du jeu de données

datasets_path="http://perso.ens-lyon.fr/lise.vaudor/grimoireStat/datasets/"
broceliande=readr::read_delim(paste0(datasets_path,'broceliande.csv'),
                              delim=';')
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   age = col_double(),
##   espece = col_character(),
##   hauteur = col_double(),
##   gui = col_double(),
##   largeur = col_double(),
##   enchantement = col_logical(),
##   fees = col_double(),
##   lutins = col_double(),
##   perlimpinpin = col_double()
## )
```

```
ggplot(broceliande, aes(x=enchantement, y=perlimpinpin))+
  geom_boxplot(fill="skyblue")+
  geom_point(data= broceliande %>%
               group_by(enchantement) %>%
               summarise(perlimpinpin=mean(perlimpinpin)),
             color="red")
```




```
#ttest avec infer
library(infer)
montest1 <- broceliande %>%
  t_test(perlimpinpin ~ enchantement,
         order=c(TRUE, FALSE))

t_obs=montest1 %>%
  select(statistic)
```

Le t-test permet de faire différents types du test de student comme test bilatéral ou unilatéral.

L'argument order me permet de dire dans quel sens j'effectue la comparaison de moyenne (ici je considère la moyenne des arbres enchantés (enchantement est TRUE) moins la moyenne des arbres non-enchantés (enchantement est FALSE) - (http://perso.ens-lyon.fr/lise.vaudor/expliquer-les-tests-statistiques-avec-le-package-infer/)

Evaluation

- Le niveau de compréhension du sujet : le sujet est un peu complex mais très clair il faut avoir des connaissances en statisitque
- L'explication des notions : les notions sont très bien expliquées, chaque ligne de code est très bien détaillé en commentaire par la suite.
- L'illustration à l'aide d'exemple : beaucoup d'illustration
- L'utilité du package : le package est très utile pour faire des statisitques
- La mise en page : la mise en page est très bien.

Conclusion

En conclusion, le sujet présenté me semble indispensable à l'utilisation de R pour faire des calculs statistiques et le fait de mettre beaucoup de documentation enrichi les explications.

Travaux comportant des aspects mathématiques explicites

1 - Regression - ZOUMANIGUI Nina

Synthèse du travail

Dans cette partie nous allons analyser le travail de ZOUMANIGUI Nina sur la régression linéaire simple et multiple. Nous avons dans le travail des définition sur la regression linéaire simple et multiple ave cles hypothèse et premisse de la regression.

Extrait commenté des parties

Je vais commenter la partie Regression multiple

Le modèle de régression linéaire multiple est l'outil statistique le plus habituellement mis en œuvre pour l'étude de données multidimensionnelles. C'est un cas particulier de modèle linéaire, il constitue la généralisation naturelle de la régression simple. On cherche avec la regresion linéaire multiple à modéliser la relation entre plus de 2 variables quantitatives. C'est très utilisé en statistique.

Evaluation

- Le niveau de compréhension du sujet : le sujet est très compréhensible.
- L'explication des notions : les notions sont très bien expliquées.
- L'illustration à l'aide d'exemple :je n'ai pas pu voir les illustration car dans son rmd les donnée ssont dans un csv stocké dans son poste de travail
- L'utilité du package : la regression est une notino très importante dans les mathématiques
- La mise en page : la mise en page est correct sauf pour les images et le code que je n'ai pas pu charger.

Conclusion

En conclusion, le sujet présenté est assez connu et important dommage qu'il n'y a pas dillustration.

2 - KMEANS - ALLAKER Maxime, BILLAUD Lucas et CHANEMOUGAM Siva

Synthèse du travail

Dans cette partie nous allons analyser le travail de ALLAKER Maxime, BILLAUD Lucas et CHANEMOUGAM Siva sur l'algorithme KMEANS. Nous avons dans le travail une définition du sujet et des termes qui sont liés à ce sujet. Puis dans une autre partie les problèmes liès à cet algo et les autres algo pouvant règler ces problèmes avec notamment l'utilisation de l'algo KMEANS ++;

Extrait commenté des parties

Je vais commenter la partie Les problèmes de k-means et une des façon de l'optimiser grâce à k-means ++

Cette partie nous explique pourquoi kmean ++ est plus optimal que kmean. L'initialisation aléatoire qui est faite dans l'algorithme kmean n'est pas efficace et est sensible aux points aberrants. L'étape d'initialisation permet d'éloignen le prochain centre le plus possible des centres déjà choisis. Les autres étapes sont les même que l'algorithme kmean.

Evaluation

- Le niveau de compréhension du sujet : le niveau de difficulté est un peu élevé mais ils ont essayé de nous expliquer les formules de la manière la plus simple.
- L'explication des notions : les notions sont très bien expliquées.
- L'illustration à l'aide d'exemple : il n'y a pas beaucoup d'illustration et peu d'exemple
- L'utilité du package : la notion semble très utile dans le domaine de machine learning non supervisé.
- La mise en page : la mise en page est correct.

Conclusion

En conclusion, le sujet présenté est assez connu et le travail fait le résume très bien, il n'y a pas beucoup d'exemple c'est peut être ce qui manque un peu.

3 - Implementation of spatial data - JUPITER Adrien et YANKO Arnaud Bruel

Synthèse du travail

Dans cette partie nous allons analyser le travail de JUPITER Adrien et YANKO Arnaud Bruel sur le sujet de l'exploitation des données spatiales. Nous avons dans un premier temps une explication sur l'interpolation en R, sur les polygones de Thiessen, puis sur la méthode d'estimation linéaire Kriging.

Extrait commenté des parties

Je vais commenter la partie Interpolation en R

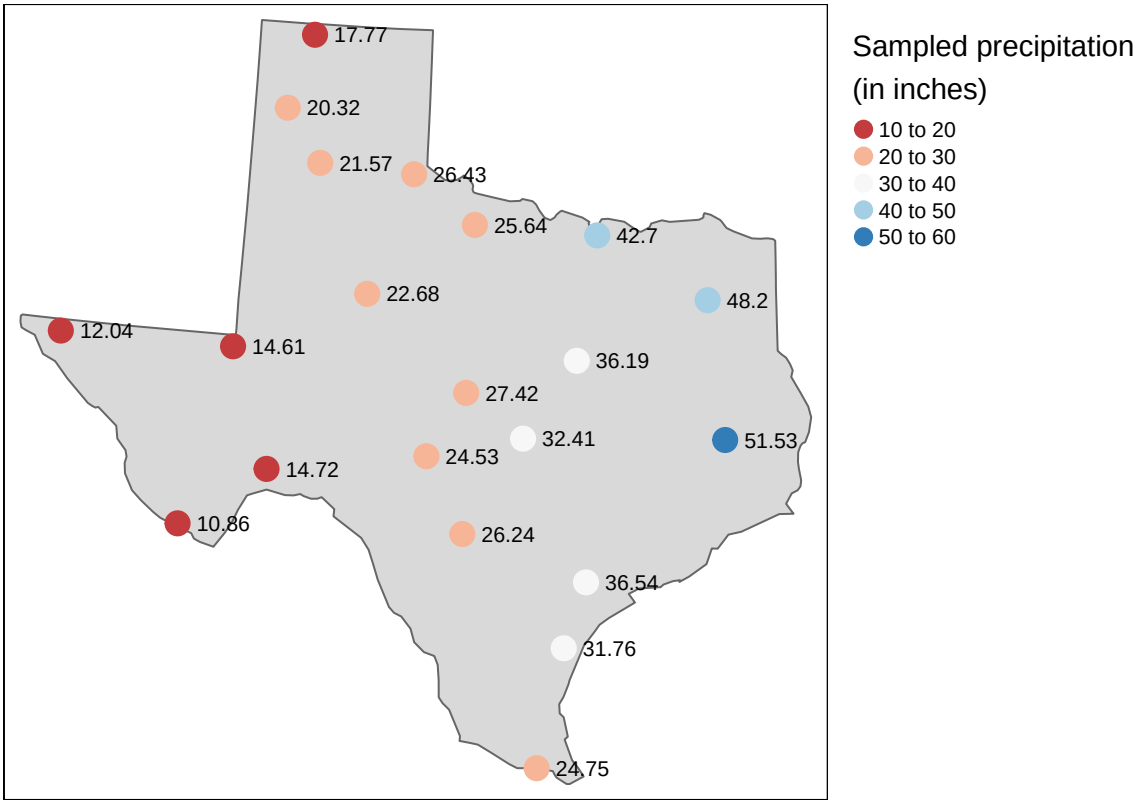
```
library(rgdal)
library(tmap)

z <- gzcon(url("http://colby.edu/~mgimond/Spatial/Data/precip.rds"))
P <- readRDS(z)

z <- gzcon(url("http://colby.edu/~mgimond/Spatial/Data/texas.rds"))
W <- readRDS(z)

P@bbox <- W@bbox

tm_shape(W) + tm_polygons() +
  tm_shape(P) +
  tm_dots(col="Precip_in", palette = "RdBu", auto.palette.mapping = FALSE,
              title="Sampled precipitation \n(in inches)", size=0.7) +
  tm_text("Precip_in", just="left", xmod=.5, size = 0.7) +
  tm_legend(legend.outside=TRUE)
```



La fonction gzcon va fournir une connexion modifiée qui encapsule une connexion existante et décompresse les lectures ou compresse les écritures via cette connexion.

La fonctino readRDS va permettre de lire la connexino qui a été faite au-dessus.

Le pakage tmap, permet de générer des cartes thématiques. La syntaxe de création est similaire à celle de ggplot2, mais adaptée aux cartes.

Cette fonction va permettre de lier les données de précipitations à l'état du Texas en remplaçant les points par des données.

Evaluation

- Le niveau de compréhension du sujet : le sujet est un peu compliqué il faut avoir des notion en géostatistique pour comprendre
- L'explication des notions : les notions sont bien expliquées mais il manque un peu plus de commnentaire
- L'illustration à l'aide d'exemple : il y a beaucoup d'illustration ce qui est indispensable à la compréhension puisque ce sont des notino concernant l'espace
- L'utilité du package : l'algorithme semble très utile notamment dans le domaine de la météorologie.
- La mise en page : la mise en page est très bien.

Conclusion

En conclusion, le sujet présenté m'était inconnu mais très interressants . Les notions sont biens expliqués à l'aide de beaucoup d'illustration.

4 - Arbres de décision - ALLIX Nicolas et LUTZ Rindra

Synthèse du travail

Dans cette partie nous allons analyser le travail de ALLIX Nicolas et LUTZ Rindra sur le sujet des arbres de décisions. Nous avons une définition théorique de l'aspect, avec notamment les notions d'arbre de classifications et d'arbre de regressions. Ensuite, ils abordent l'histoire de ce sujet puis nous expliquent en détail l'aspect mathématique et comment construire un arbre de décision. On peut également trouver des exemples et les avantages et limites du sujet.

Extrait commenté des parties

Je vais commenter la partie Construction d'un arbre de décision.

Cette partie nous explique comment construire un arbre de décision. Il faut d'abord débuter la construction de l'arbre en partant de la gauche avec la décision à prendre. Puis, pour chaque option possible, on trace un trait avec le nom à côté. A la fin de ce trait, si une nouvelle décision est à prendre, on ajoute de nouveau une décision. Cet outil va finalement permettre d'obtenir des résultats prévisionnels chiffrés selon les options retenues. C'est une aide très importante pour alimenter la prise de décision.

Evaluation

- Le niveau de compréhension du sujet : ce n'est pas trop compliqué de comprendre le sujet, les éléments sont claires
- L'explication des notions : les notions sont très bien expliquées.
- L'illustration à l'aide d'exemple : il y a une partie entière dédiée à l'exemple avec des images pour bien comprendre, c'est très bien.
- L'utilité du package : la notion semble très utile notamment pour la prise de décisions.
- La mise en page : la mise en page est correct.

Conclusion

En conclusion, le sujet présenté est assez connu et le travail fait le résume très bien, les exemples m'ont très bien aidé à comprendre le sujet car c'est très bien illustré.

5 - Algorithme génétique - COMLAN Florine et HOUNTONDJI Ramya

Synthèse du travail

Dans cette partie nous allons analyser le travail de COMLAN Florine et HOUNTONDJI Ramya sur le sujet de l'algorithme génétique. Nous avons dans un premier temps l'histoire de l'agorithme génétique, puis les fondamentaux. Une troisième partie très intéressante sur le détail de l'algorithme. Elles nous expliquent ensuite dans quels domaines sont appliqués cet algortihme puis les avantages et limites et ce qui le différencie des autres algorithmes.

Extrait commenté des parties

Je vais commenter dans la partie Description détaillé, la fonction fitness.

La fonction fitness évalue dans quelle mesure une solution donnée est proche de la solution optimale du problème souhaité. Elle détermine dans quelle mesure une solution est adaptée.

En entrée nous avons une solution candidate au problème et en sortie la mesure dans laquelle la solution est bonne par rapport au problème.

Dans les algorithmes génétiques, chaque solution est généralement représentée par une chaîne de nombres binaires, connue sous le nom de chromosome. Nous devons tester ces solutions et trouver le meilleur ensemble de solutions pour résoudre un problème donné. Chaque solution doit donc recevoir un score, pour indiquer dans quelle mesure elle est proche de la spécification globale de la solution souhaitée. Ce score est généré en appliquant la fonction fitness au test, ou les résultats obtenus à partir de la solution testée.

Le calcul fait doit être très rapide car il y a plusieurs tests à faire avant de trouver une bonne solution.

Evaluation

- Le niveau de compréhension du sujet : le sujet est un peu compliqué mais elles ont réussi à l'expliquer très clairement
- L'explication des notions : les notions sont très bien expliquées.
- L'illustration à l'aide d'exemple : il y a beaucoup d'illustration ce qui est indispensable à la compréhension
- L'utilité du package : l'algorithme semble très utile notamment concernant la partie optimisation
- La mise en page : la mise en page est très bien.

Conclusion

En conclusion, le sujet présenté m'était inconnu j'espère pouvoir l'utiliser. Les notions sont très biens expliqués à l'aide de beaucoup d'exemple et la partie domaine d'application est très intérresantes, j'ai pu voir concrètement voir à quel moment cet algo est utilisé.

Travaux auxquels j'ai participé

1 - PLOTLY

J'ai travaillé sur le package R Plotly avec Imen DERROUICHE, avec la mise en place d'un fichier assez complet, nous sommes allé des fonctinos les plus basique à celles un peu plus complexes.

Nous avons traiter les des graphiques

Nous avons définit chaque partie avec des définitions, illustré chaque partie avec des exemples, expliqué, pour chaque ligne de code, à quoi servaient les paramètres et cité nos sources.

Nous avons eu parfois des difficultés avec certains paramètres car il fallait installer du code open source à part.

Comme Plotly est dynamique nous n'avons pas pu l'exporter en pdf ce qui a pu être dérangeant pour nos camarades.

Globalement nous sommes plûtot satifait de notre travail car nous avons bien expliqué le sujet avec beaucoup d'exemple et pour tous les niveaux.

2 - Automating biomedical data science through tree-based pipeline optimization

J'ai travaillé sur le package TPOT dans le biomédical avec Imen DERROUICHE et Marion DANYACH, avec l'explication de la conception des pipelines.

Nous avons traiter ce sujet qui est dans le domaine du machine learning et qui permet d'aider à donner des idées sur la manière de résoudre un problème d'apprentissage particulier en explorant des configurations de pipeline.

Nous avons définit TPOT et nous avons égalemetn fait le parallèle avec les algo génétiques.

Il est très compliqué de trouver des exemples simple de code nous avons donc expliquer à l'aide d'illustration au lieu de code.

Globalement nous sommes plûtot satifait de notre travail car nous avons inintié les lecteurs à TPOT qui va remplacer l'Homme dans des phases d'apprenttissage très lourd dans le domaine du machine learning.