

Référentiel de certification

Intitulé de la certification : R niveau 1 - Utilisateur – Analyse de données

Objectifs de la certification : Le titulaire de la certification est capable, dans une démarche de data science menée par son organisation, d'intégrer et de traiter des données, et de produire un rapport en utilisant de manière appropriée le langage de programmation R. Le titulaire de la certification est capable de mettre en place un processus de traitement des données sous R qui est reproductible avec différents jeux de données.

Prérequis : Avoir déjà rédigé une ligne de code (dont formule Excel). Connaître les concepts de paramètres de position (moyenne, médiane) et de dispersion (variance, écart-type).

Glossaire :

- Package : lignes de codes préexistantes correspondant à une ou plusieurs fonctionnalités.
- Environnement : interface utilisateur de saisie des lignes de codes et de téléchargement des packages.
- Reproductibilité : exécuter sans modifications suite à une mise à jour des données. En langage de développement, l'expression est « les lignes de codes peuvent être rejouées ».

ACTIVITE et TACHES	COMPETENCES ASSOCIEES AUX ACTIVITES ET TACHES	MODALITES D'EVALUATION	CRITERES D'EVALUATION
<p>A1. Importation d'un jeu de donnée dans l'environnement R</p> <p>Configuration de l'environnement de travail dans le logiciel R</p> <p>Chargement des packages adaptés</p> <p>Importation du fichier de données</p> <p>Vérification de la qualité de l'import</p> <p>Typage des données (ce qu'est chaque colonne : date, numérique, texte...)</p>	<p>Importer des données dans un environnement R adapté, en utilisant le(s) package(s) adéquat(s) et les lignes de code adaptées aux données, afin de vérifier leur qualité et réaliser leur mise en forme.</p>	<p>Cas pratique : Sur la base d'un jeu de données, contenant des erreurs, le candidat doit :</p> <ul style="list-style-type: none"> - Importer les données ; - Corriger les erreurs ; - Décrire et compiler les données dans un rapport mis en forme selon des spécifications précises. <p><i>Critère général de forme : Le candidat respecte les bonnes pratiques d'écriture et de commentaires des lignes de code.</i></p>	<p>Le projet existe dans l'interface R et les packages nécessaires au projet sont chargés.</p> <p>Le candidat respecte les bonnes pratiques de configuration de son environnement logiciel, permettant son utilisation par un autre utilisateur.</p> <p>Les données sont visualisables.</p> <p>Les données sont correctement formatées / typées</p> <p>Les lignes de code d'import et de typage peuvent être rejouées.</p>

<p>A2. Visualisation d'un jeu de donnée dans l'environnement R</p> <p>Production de graphiques et tableaux à partir des données</p> <p>Description statistique des données : calcul des paramètres descriptifs pertinents (effectifs, moyennes, médianes, écarts-types...)</p>	<p>Visualiser des données importées dans l'environnement R sous forme de graphiques, tableaux et d'indicateurs statistiques adaptés à leur nature (moyennes, médianes et écarts-types...)</p>		<p>Existence des statistiques descriptives sur les variables visées.</p> <p>Le candidat est capable de visualiser les données sous forme de tableau ou de graphique.</p> <p>Les lignes de code de description et de visualisation des données peuvent être rejouées.</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p>A3. Nettoyage d'un jeu de donnée dans l'environnement R</p> <p>Interprétation des indicateurs statistiques (moyennes, médianes, écarts-types)</p> <p>Identification des erreurs de saisie ou de collecte</p> <p>Correction des erreurs par codage en langage R</p> <p>Visualisation des données à l'aide des lignes de codes appropriés</p> <p>Recherche d'erreurs non détectées lors des vérifications initiales</p> <p>Corrections des nouvelles erreurs détectées</p> <p>Transformation des données pertinentes pour rendre le jeu de données complètement exploitable</p>	<p>Interpréter les indicateurs statistiques produits dans l'environnement R pour identifier des erreurs de saisie ou de collecte et les corriger à l'aide des lignes de code appropriées, afin de rendre le jeu de données complètement exploitable</p>		<p>Le candidat a détecté toutes les erreurs.</p> <p>Le candidat a corrigé toutes les erreurs.</p> <p>Les lignes de code de correction des données peuvent être rejouées.</p> <p>Le jeu de données est complètement exploitable.</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p><i>A4. Production d'un rapport à partir d'un jeu de donnée dans l'environnement R</i></p> <p>Choix des variables et indicateurs statistiques selon la commande ou l'intérêt des données</p> <p>Sélection des formats de visualisation des variables et indicateurs statistiques</p> <p>Production des tableaux et graphiques attendus ou pertinents pour la compréhension et l'exploitation des données</p> <p>Mise en forme des tableaux et graphiques (titre, couleur, traits, proportions...)</p>	<p>Sélectionner des variables et d'indicateurs statistiques pertinents en fonction des données ou de la commande, et produire des tableaux et des graphiques mis en forme, permettant de comprendre et interpréter les données, en rédigeant les lignes de codes adaptées, afin de produire un rapport pour un tiers.</p>		<p>Les variables demandées sont représentées.</p> <p>Les tableaux et graphiques demandés sont réalisés.</p> <p>Les critères de mise en forme demandés sont respectés.</p> <p>Les lignes de codes sont agencées de manière logique.</p> <p>Les lignes de code de production des tableaux et graphiques peuvent être rejouées : le rapport est reproductible.</p>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------