

Sommaire

Remerciements	i
Prologue	iii
1 Python en sciences humaines et sociales	1
1.1 Pourquoi programmer et pourquoi Python ?	1
1.2 Les SHS et la programmation	2
1.3 Quelle place pour la programmation en SHS ?	4
1.4 Les spécificités de Python par rapport à d'autres langages	6
1.5 Installation de Python	10
1.6 Résumé du chapitre	12
2 Se familiariser à Python	15
2.1 Penser comme un ordinateur	15
2.2 Quelques conseils pour débiter	17
2.3 Où faire du Python ?	18
2.3.1 Différentes possibilités pour parler à l'ordinateur	18
2.3.2 Programmer dans une console	19
2.3.3 Dans votre navigateur internet	21
2.3.4 Les environnements de développement intégrés	21
2.4 Écrire notre première ligne de code	22
2.4.1 Utiliser le Notebook Jupyter	22
2.4.2 Cellules de codes dans ce livre	26

2.4.3	Écrire un script dans un fichier	26
2.4.4	Écrire un script un peu plus long	28
2.5	Faire des erreurs	29
2.5.1	Ne paniquez pas	29
2.5.2	Trouver de l'aide	31
2.6	Créer et manipuler des variables	32
2.6.1	Qu'est-ce qu'une variable ?	32
2.6.2	Le type des variables	35
2.6.3	Les textes ou chaînes de caractères	36
2.6.4	Les ensembles : listes et dictionnaires	38
2.6.5	Conditions et comparaisons	43
2.6.6	Encore plus de types	44
2.6.7	#ViveLesCommentaires	45
2.7	Les blocs structurants : conditions, boucles, fonctions	46
2.7.1	Si [condition] alors [faire]	47
2.7.2	Les boucles pour répéter les opérations	49
2.7.3	Construire ses fonctions	51
2.7.4	Une multitude de fonctions	55
2.7.5	Un mot sur l'espace des noms	57
2.8	Aller plus loin pour gérer les erreurs	59
2.9	Synthèse du chapitre	60
3	Les bibliothèques	61
3.1	Qu'est-ce qu'une bibliothèque ?	61
3.2	Un écosystème	62
3.3	Utiliser des bibliothèques	65
3.3.1	Charger un module déjà installé	65
3.3.2	Installer une nouvelle bibliothèque	67
3.3.3	Quelques problèmes que vous pouvez rencontrer	68
3.4	Mise en pratique	69
3.4.1	Définir ses besoins, chercher l'information	69

3.4.2	Utiliser une nouvelle bibliothèque	71
3.5	Les évolutions des bibliothèques	74
3.6	Synthèse du chapitre	75
4	Manipuler les données	77
4.1	Pourquoi manipuler des données ?	77
4.2	Charger et sauvegarder des fichiers	78
4.2.1	Quelques mots sur les fichiers et leurs formats	79
4.2.2	Ouvrir un fichier	80
4.2.3	Lire des informations dans un fichier	81
4.2.4	Écrire dans un fichier texte	82
4.2.5	Ouvrir d'autres fichiers	84
4.3	Manipuler les ensembles d'éléments	85
4.3.1	Rappel sur les listes	85
4.3.2	Les listes emboîtées	87
4.4	Traiter du texte	89
4.4.1	Il faut qu'on parle... d'encodage	89
4.4.2	Mettre en forme le texte	90
4.4.3	Intégrer des variables	91
4.4.4	Traiter des données réelles	92
4.4.5	Introduction aux expressions régulières	95
4.5	Complément : attention aux attributions	98
4.6	Synthèse du chapitre	99
5	Traitement de données avec <i>Pandas</i>	101
5.1	La bibliothèque <i>Pandas</i>	101
5.2	Charger des données à partir d'un fichier	103
5.3	Le format des données <i>Pandas</i>	104
5.4	Modifier un tableau <i>Pandas</i>	107
5.4.1	Sélectionner des colonnes	107
5.4.2	Manipuler l'index	108
5.4.3	Renommer les colonnes	109

5.4.4	Créer une nouvelle colonne	110
5.4.5	Modifier une case du tableau	110
5.4.6	Supprimer des lignes ou des colonnes	111
5.4.7	Répéter une opération sur un tableau	111
5.4.8	Gérer les valeurs manquantes	116
5.5	Filtrer des informations	116
5.6	Relier des informations de tableaux différents	118
5.6.1	Passer par un dictionnaire	119
5.6.2	Joindre des tableaux	121
5.7	Créer un tableau <i>Pandas</i>	122
5.8	Sauvegarder ses données	123
5.9	Synthèse du chapitre	124
6	Statistiques descriptives et inférentielles	125
6.1	Passer de la programmation aux statistiques	125
6.2	Construire et nettoyer ses tableaux	127
6.2.1	Le tableau de données	127
6.2.2	Vérifier la qualité des données	128
6.2.3	Sélectionner les données utiles	129
6.3	Recoder les variables	130
6.3.1	Construire des indicateurs	130
6.3.2	Recodage des variables numériques en intervalles	134
6.4	Analyse univariée des données	135
6.4.1	Le tri à plat des variables	135
6.4.2	Les indicateurs de tendance centrale et de dispersion	138
6.5	Relation dans les données	144
6.5.1	Construire des tableaux croisés	144
6.5.2	Les corrélations	148
6.5.3	Grouper des catégories	149
6.6	Réaliser des tests statistiques	150
6.6.1	Quelques rappels	150

6.6.2	Intervalle de confiance	153
6.6.3	Significativité d'un tableau croisé	155
6.6.4	Le V de Cramer	156
6.6.5	Différence entre deux groupes	157
6.7	Synthèse du chapitre	159
7	Visualiser	161
7.1	Pourquoi programmer des visualisations?	161
7.2	Petit tour des lieux	163
7.2.1	Représenter les données avec <i>Pandas</i>	164
7.2.2	<i>Matplotlib</i> , une bibliothèque de bas niveau	165
7.2.3	<i>Seaborn</i> , une bibliothèque de haut niveau	166
7.3	La grammaire des visualisations de <i>Matplotlib</i>	167
7.3.1	Quelques éléments généraux	167
7.3.2	Définir une figure	168
7.3.3	Les différents types de figures	171
7.3.4	Tracer plusieurs figures	178
7.3.5	Couleurs et légendes	184
7.3.6	Sauvegarder une figure	185
7.3.7	Affiner une visualisation	186
7.4	Visualiser avec <i>Pandas</i>	188
7.4.1	Les différents types de visualisation	188
7.4.2	Combiner <i>Pandas</i> et <i>Matplotlib</i>	188
7.4.3	Combiner tableaux et visualisations	189
7.5	Visualisations avancées avec <i>Seaborn</i>	191
7.5.1	Quelques visualisations utiles	191
7.5.2	Stylisation des figures avec les thèmes	195
7.6	Aller plus loin dans la maîtrise du visuel	195
7.6.1	Séparer exploration et finalisation	195
7.6.2	Exemple d'une figure complexe : le <i>bubble chart</i>	196
7.6.3	Autres visualisations	198

7.7	Synthèse du chapitre	199
8	Statistiques avancées	201
8.1	Les traitements avancés en Python	201
8.2	Les analyses factorielles	202
8.2.1	Une grande famille	202
8.2.2	Analyse en composantes principales	203
8.2.3	Analyse des correspondances multiples	211
8.2.4	Analyse factorielle des correspondances	214
8.3	Les modèles de régression	216
8.3.1	Quelques éléments sur la modélisation	216
8.3.2	Régressions linéaires	217
8.3.3	Régressions logistiques	219
8.4	Les classifications	225
8.4.1	Quelques éléments généraux	225
8.4.2	Partitionnement en k-moyennes	226
8.4.3	Classification hiérarchique ascendante	228
8.5	Les modèles d'apprentissage automatique	232
8.5.1	Quelques éléments généraux	232
8.5.2	Régression logistique pour prédire	233
8.6	Synthèse du chapitre	235
9	Usages avancés	237
9.1	Récupérer des données complexes	237
9.1.1	Transformer du matériau qualitatif en données	237
9.1.2	Récupérer des données sur internet	244
9.1.3	Utiliser des API : <i>Twitter</i> , <i>Google</i> et les autres	250
9.2	Analyse de corpus textuels	258
9.2.1	Traiter des textes pour en dégager de l'information	258
9.2.2	L'information relationnelle dans les réseaux	273
9.3	Visualiser des données géographiques	281
9.3.1	La construction de cartes	281

9.3.2	Introduction à <i>GeoPandas</i>	282
9.3.3	La carte des populations françaises	286
9.4	Synthèse du chapitre	290
10	Organiser son code et le partager	291
10.1	De la ligne au logiciel	291
10.2	Organiser son script dans le Notebook	293
10.3	Retravailler une visualisation	294
10.3.1	Identifier les éléments génériques dans un code	294
10.3.2	Aller vers une fonction	297
10.3.3	Un fichier à soi : créer un module dédié	300
10.4	Créer une bibliothèque	302
10.5	Collaborer autour du code	304
10.6	Synthèse du chapitre	304
	Conclusion	305
A	Annexe : les coulisses du code	309
A.1	Règles à avoir en tête	309
A.2	Écrire un code pas-à-pas	310
A.2.1	Identifier un objectif	311
A.2.2	Procéder par étapes	311
A.2.3	Généraliser vers une fonction	318
A.2.4	Documentation	320
A.3	Synthèse	320
B	Ressources utiles pour l'apprentissage	321
	Bibliographie	323
	Index des fonctions	327
	Index	331