
Coding in GIS

Nils Ratnaweera

Sep 08, 2020

CODING IN GIS I

1	Einleitung zu diesem Block	3
2	Input: Primitive Datentypen	5
2.1	Boolean	5
2.2	String	5
2.3	Integer	6
2.4	Float	6
2.5	Weitere Datentypen	6
3	Input: Zusammengesetzte Datentypen	7
3.1	List	7
3.2	Dict	8
3.3	DataFrame	8
3.4	Tuples	9
3.5	Sets	10
4	Übung: Lists	11
4.1	Übung 1: Variablen erstellen	11
4.2	Übung 2: Lists	12
4.3	Übung 3: Elemente aus Liste ansprechen	12
4.4	Übung 4: Liste ergänzen	12
5	Übung: Dictionaries	13
5.1	Übung 1: Dictionary	14
5.2	Übung 2: Elemente aus Dictionary ansprechen	14
5.3	Übung 3: Key ergänzen	14
5.4	Input	15
5.5	Übung 4: Dictionary mit List	15
5.6	Übung 5: Functions	15
6	Input: Python Modules	17
6.1	Erweiterung installieren	17
6.2	Erweiterung laden	17
6.3	Erweiterung verwenden	18
6.4	Modul mit Alias importieren	18
6.5	Einzelne <i>Function</i> importieren	18
6.6	Alle <i>Functions</i> importieren	18
7	Übung: Tabellarische Daten	21
7.1	Übung 1: von einer <i>Dictionary</i> zu einer <i>DataFrame</i>	21
7.2	Übung 2: <i>DataFrame</i> in csv umwandeln	21

7.3	Übung 3: CSV als <i>DataFrame</i> importieren	21
7.4	Übung 4: Koordinaten räumlich darstellen	22
7.5	Übung 5: Einzelne Spalte selektieren	22
7.6	Übung 6: Neue Spalte erstellen	22
8	Einleitung zu diesem Block	25
9	Input: <i>Function</i> Basics	27
10	Übung: <i>Functions</i>	29
10.1	Übung 1: Erste <i>Function</i> erstellen	29
10.2	Übung 2: <i>Function</i> erweitern	29
10.3	Übung 3: Default-Werte festlegen	30
11	Input: <i>Function</i> Advanced	31
11.1	<i>Functions</i> mit Default Werten	31
11.2	Reihenfolge der Argumente	31
11.3	Globale und Lokale Variablen	32
12	Übung: Einen Einzelpunkt zufällig verschieben	33
12.1	Übung 1: Zufallswerte generieren	34
12.2	Übung 2: Zufallswerte addieren	35
12.3	Übung 3: Arbeitsschritte in eine <i>Function</i> verwandeln	35
12.4	Übung 4: Output visualisieren	35
13	Übung Punkte einer <i>DataFrame</i> zufällig verschieben	37
13.1	Übung 1: Alle Zeckenstiche zufällig verschieben	38
13.2	Übung 2: Neue <i>GeoDataFrame</i> mit simulierten Punkten erstellen	38
13.3	Übung 4: Mehrere <i>DataFrames</i> visualisieren	39
14	Einleitung zu diesem Block	41
15	Übung: For Loops (Teil I)	43
15.1	Übung 1: Erste For-Loop erstellen	43
15.2	Übung 2: For-Loop mit <code>range()</code>	44
16	Übung: For Loops (Teil II)	45
16.1	Übung 1: Output aus For-Loop speichern	46
17	Zeckenstich Simulation mit Loop	47
17.1	Übung 1: Mit For-Loop zeckenstiche mehrfach verschieben	47
17.2	Übung 2: <i>DataFrames</i> aus Simulation zusammenführen	48
17.3	Übung 3: Simulierte Daten visualisieren	49
18	Übung: GIS in Python	51
18.1	Übung 1: <i>DataFrame</i> zu <i>GeoDataFrame</i>	51
18.2	Übung 2: Koordinatensystem festlegen	52
18.3	Übung 3: Zeckenstiche als Shapefile exportieren	53
19	Waldanteil berechnen	55
19.1	Übung 1: Wald oder nicht Wald?	55
19.2	Übung 2: Anteil der Punkte im Wald	57
19.3	Übung 3: Mittelwerte Visualisieren	57

Dieser kurze Kurs ist Bestandteil des übergreifenden Moduls “[Angewandte Geoinformatik](#)” der Zürcher Hochschule für Angewandte Wissenschaften (ZHAW). Er soll einen Einstieg in die Programmierwelt von Python bieten und spezifisch zeigen wie man räumliche Fragestellungen mit frei verfügbaren Programmen lösen kann. Die Voraussetzung für diesen Kurs ist eine Offenheit neue Tools und Ansätze kennen zu lernen sowie die Bereitschaft für lösungsorientiertes arbeiten und etwas Hartnäckigkeit.

Das github-repo ist hier [ratnanil/codingingis](https://github.com/ratnanil/codingingis).

Dieses Buch ist sowohl online als auch pdf version verfügbar:

- online: <https://ratnanil.github.io/codingingis>
 - pdf: <https://github.com/ratnanil/codingingis/raw/master/codingingis.pdf>
-

Typographie

Noch ein kurzer Hinweis zur Typografie dieses Dokumentes:

- Wenn sich im Fliesstext (Python- oder R-) Code befindet, wird er in dieser `Festschriftart` dargestellt
- Englische Begriffe, deren Übersetzung eher verwirrend als nützlich wären, werden *in dieser Weise* hervorgehoben
- Alleinstehende Codezeilen werden folgendermassen dargestellt:

```
print("Coding in GIS!")
```

- Da viele von euch bereits Erfahrung in R haben, stelle ich immer wieder den Bezug zu dieser Programmiersprache her. Diese Verweise sind folgendermassen gekennzeichnet:

Für R Nutzer

- hier wird ein Bezug zu R gemacht
-

In der Online Version dieses Buches werdet ihr häufig am rechten Rand einen Button mit der Aufschrift “Click to show” antreffen (in der PDF Version des Buches ist dieses interaktive Element nicht verfügbar). Im Kontext von Aufgaben seht ihr mit einem Klick auf das Plus-Symbol die Musterlösung der entsprechenden Aufgabe. Ich empfehle euch, diesen Button nur sehr zögerlich zu verwenden! Wenn ihr euch durch die Musterlösungen klickt werdet ihr sehr wenig aus diesem Modul mitnehmen.

```
# hier steht die Musterlösung der Aufgabe. Erst betätigen wenn ihr  
# vergeblich versucht habt, die Aufgabe selbst zu lösen!  
print("Don't peek!")
```


EINLEITUNG ZU DIESEM BLOCK

Übungsziele

- JupyterLabs aufstarten, kennenlernen und bei Bedarf personalisieren
 - Python kennen lernen, erste Interaktionen
 - Die wichtigsten Datentypen in Python kennen lernen (bool, text, integer, float, list, dictionary, dataframe)
 - Pandas DataFrames kennen lernen und einfache Manipulationen durchführen
-

INPUT: PRIMITIVE DATENTYPEN

Bei primitiven Datentypen handelt es sich um die einfachste Datenstruktur. Sie werden deshalb auch “atomare Datentypen” genannt: Alle komplexeren Datentypen (Tabellarische Daten, Bilder, Geodaten) basieren auf diesen einfachen Strukturen.

Für R Nutzer

- In R gibt es die gleichen Primitiven Datentypen wie in Python: Boolean, String, Integer, Float. Die Bezeichnung ist jedoch leicht anders (Logical, Character, Integer, Numeric, Double)
- In R können Variablen genau gleich zugewiesen werden wie in Python (mit dem = Symbol). In R gibt es zudem noch die Möglichkeit, Variablen mit <- zu zuweisen, in Python jedoch nicht

2.1 Boolean

Hierbei handelt es sich um den einfachsten Datentyp. Er beinhaltet nur zwei Zustände: Wahr oder Falsch. In Python werden diese mit `True` oder `False` definiert. Beispielsweise sind das Antworten auf geschlossene Fragen.

```
regen = True
```

```
sonne = False
```

```
type(sonne)
```

```
bool
```

2.2 String

In sogenannten *Strings* werden Textinformationen gespeichert. Beispielsweise können das die Namen von Ortschaften sein.

```
stadt = "Bern"  
land = "Schweiz"
```

```
type(stadt)
```

```
str
```

Strings können mit + miteinander verbunden werden

```
stadt + " ist die Hauptstadt der " + land
```

```
'Bern ist die Hauptstadt der Schweiz'
```

2.3 Integer

In Integerwerten werden ganzzahlige Werte gespeichert, beispielsweise die Anzahl Einwohner einer Stadt.

```
bern_einwohner = 133115
```

```
type(bern_einwohner)
```

```
int
```

2.4 Float

Als Float werden Zahlen mit Nachkommastellen gespeichert, wie zum Beispiel die Temperatur in Grad Celsius.

```
bern_temp = 22.5
```

```
type(bern_temp)
```

```
float
```

2.5 Weitere Datentypen

Es gibt noch weitere Datentypen, auf die wir an dieser Stelle jedoch nicht eingehen.

INPUT: ZUSAMMENGESETZTE DATENTYPEN

Basierend auf den “atomaren” oder primitiven Datentypen existieren komplexere Datenstrukturen, die aber an sich immer noch sehr einfach sind.

3.1 List

Eine List:

- speichert die Reihenfolge, in der die Werte eingegeben werden
- kann unterschiedliche Datentypen enthalten
- wird mit eckigen Klammern ([und]) erstellt

```
hexerei = [3,1,2]
```

Der erste Wert wird in Python mit 0 aufgerufen:

```
hexerei[0]
```

```
3
```

```
type(hexerei)
```

```
list
```

In einer List können verschiedene Datentypen enthalten sein, auch weitere, verschachtelte Lists

```
chaos = [23, "ja", [1,2,3]]
```

```
# Der Inhalt vom ersten Wert ist vom Typ "Int"  
type(chaos[0])
```

```
int
```

```
# Der Inhalt vom dritten Wert ist vom Typ "List"  
type(chaos[2])
```

```
list
```

Für R Nutzer

Eine Python List entspricht eines R Vectors

3.2 Dict

Eine Dict:

- speichert die Reihenfolge, in der die Werte eingegeben werden **nicht**
- kann unterschiedliche Datentypen enthalten
- wird mit geschweiften Klammern (`{` und `}`) erstellt
- speichert für jeden Wert («Value») einen Schlüssel («Key»), mit dem er abgerufen werden kann:

```
dict_name = {"Key" : "Value"}
```

```
menue = {"Vorspeise": "Suppe",  
         "Hauptspeise": "Gratin",  
         "Dessert": "Eis"}
```

```
menue["Vorspeise"]
```

```
'Suppe'
```

Auch hier sind verschiedene Datentypen möglich:

```
menue = {"Vorspeise": "Suppe",  
         "Hauptspeise": ["Gratin", "Spinat", "Salat"],  
         "Dessert": "Himbeerglacé",  
         "Preis": 50}
```

```
type(menue)
```

```
dict
```

3.3 DataFrame

Bei den bisherigen Datentypen handelte es sich um Strukturen, die in der Standardinstallation von Python enthalten sind. Tabellarische Strukturen sind in Python, im Gegensatz zu R, nicht standardmäßig vorhanden. Dazu brauchen wir eine Erweiterung zu Python: Was es sich damit auf sich hat und wie diese installiert wird erfahren wir später. An dieser Stelle möchte ich nur die Struktur `DataFrame` vorstellen. `DataFrames` sind:

- Tabellarische Daten
- Ein Spezialfall einer Dict:
 - values sind Lists von gleicher Länge
 - alle Werte in einer List bestehen aus dem gleichen Datentyp
- jeder key ist ein Spaltenname

```
import pandas as pd
```

```
menue2 = {"Typ": ["Vorspeise", "Hauptspeise", "Dessert"],
          "Beschreibung": ["Suppe", "Gratin mit Spinat", "Himbeerglacé"],
          "Preis": [7.50, 32.0, 10.50]}
```

```
menue2
```

```
{'Typ': ['Vorspeise', 'Hauptspeise', 'Dessert'],
 'Beschreibung': ['Suppe', 'Gratin mit Spinat', 'Himbeerglacé'],
 'Preis': [7.5, 32.0, 10.5]}
```

```
type(menue2)
```

```
dict
```

```
menue_df = pd.DataFrame(menue2)
```

```
menue_df
```

	Typ	Beschreibung	Preis
0	Vorspeise	Suppe	7.5
1	Hauptspeise	Gratin mit Spinat	32.0
2	Dessert	Himbeerglacé	10.5

```
type(menue_df)
```

```
pandas.core.frame.DataFrame
```

3.4 Tuples

Tuples sind ähnlich wie Lists, nur können sie nachträglich nicht verändert werden. Das heisst, es ist nach der Erstellung keine Ergänzung von neuen Werten oder Löschung von bestehenden Werten möglich. Tuples werden mit runden Klammern ((und)) erstellt.

Dieser Datentyp spielt in unserem Kurs *Coding in GIS* eine untergeordnete Rolle.

```
gewinner = (1,2,3)
```

```
type(gewinner)
```

```
tuple
```

3.5 Sets

Sets sind ähnlich wie Dicts, jedoch kommt jeder Wert nur 1x vor. Zudem verfügen Sets nicht über keys und values. Sie werden mit geschweiften Klammern ({ und }) erstellt.

Dieser Datentyp spielt in unserem Kurs *Coding in GIS* eine untergeordnete Rolle.

```
hexerei2 = {1, 2, 3, 1, 2, 2, 1, 3}
```

```
hexerei2
```

```
{1, 2, 3}
```

```
type(hexerei2)
```

```
set
```

ÜBUNG: LISTS

4.1 Übung 1: Variablen erstellen

Wir beginnen damit, einfache Datentypen wie Zahlen, Text, boolsche Variablen (siehe Vorlesungsfolien) sogenannten Variablen zu zuweisen. Führe die folgenden Schritte aus und schau dir nach jedem Schritt die erstellten Variablen jeweils im Variable explorer von Spyder an

1. Erstelle eine Variabel `vorname` mit deinem Vornamen
2. Erstelle eine zweite Variabel `nachname` mit deinem Nachnamen
3. Was sind `vorname` und `nachname` für Datentypen?
4. Klebe die beiden Variablen zusammen, ohne den Output zu speichern
5. Erstelle eine Variabel `groesse` mit deiner Körpergröße in Zentimeter. Was ist das für ein Datentyp?
6. Ermittle deine Größe in Meter auf der Basis von `groesse`. Was ist das für ein Datentyp?
7. Erstelle eine boolsche Variable `blond` und setze sie auf `True` wenn diese Eigenschaft auf dich zutrifft und `False` falls nicht.

```
# Musterlösung

vorname = "Guido"
nachname = "van Rossum"

type(vorname) # es handelt sich um den Datentyp "str", also String (Text)
vorname+nachname

groesse = 184
type(groesse) # es handelt sich hierbei um den Datentyp "integer"

groesse_meter = groesse/100
type(groesse_meter) # es handelt sich um den Datentyp "float"

blond = False
```

4.2 Übung 2: Lists

1. Erstelle eine Variable `vornamen` bestehend aus einer *List* mit 3 Vornamen
2. Erstelle eine zweite Variable `nachnamen` bestehend aus einer *List* mit 3 Nachnamen
3. Erstelle eine Variable `groessen` bestehend aus einer *List* mit 3 Grössenangaben in Zentimeter.

```
# Musterlösung

vornamen = ["Christopher", "Henning", "Severin"]
nachnamen = ["Annen", "May", "Kantereit"]

groessen = [174, 182, 162]
```

4.3 Übung 3: Elemente aus Liste ansprechen

Wie erhältst du den ersten Eintrag in der Variable `vornamen`? Wie erhältst du den letzten Eintrag? Tipp: nutze dazu `[` und `]` sowie eine Zahl.

Für R Nutzer

In R werden Elemente aus Vectors auf die gleiche Weise extrahiert. Nur ist in R das erste Element die Nummer 1, Python beginnt bei 0

4.4 Übung 4: Liste ergänzen

Listen können durch der Method `append` ergänzt werden.

```
vornamen.append("Malte")
```

Ergänze in die Listen `vornamen`, `nachnamen` und `groessen` durch je einen Eintrag.

```
# Musterlösung

nachnamen.append("Huck")

groessen.append(177)
```


ÜBUNG: DICTIONARIES

Ähnlich wie eine List, ist eine Dictionary ein Behälter wo mehrere Elemente abgespeichert werden können. Wie bei einem Wörterbuch bekommt jedes Element ein “Schlüsselwort”, mit dem man den Eintrag finden kann. Unter dem Eintrag “trump” findet man im Langenscheidt Wörterbuch (1977) die Erklärung “erdichten, schwindeln, sich aus den Fingern saugen”.



In Python würde man diese *Dictionary* folgendermassen erstellen:

```
langenscheidt = {"trump": "erdichten- schwindeln- sich aus den Fingern saugen"}
```

Schlüssel (von nun an mit *Key* bezeichnet) des Eintrages lautet “trump” und der dazugehörige Wert (*Value*) “erdichten- schwindeln- aus den Fingern saugen”. Beachte die geschweiften Klammern ({ und }) bei der Erstellung einer Dictionary.

Eine *Dictionary* besteht aber meistens nicht aus einem, sondern aus mehreren Einträgen: Diese werden Kommage- trennt aufgeführt.

```
langenscheidt = {"trump": "erdichten- schwindeln- sich aus den Fingern saugen",  
↪ "trumpery": "Plunder- Ramsch- Schund"}
```

Der Clou der *Dictionary* ist, dass man nun einen Eintrag mittels dem *Key* aufrufen kann. Wenn wir also nun wissen wollen was “trump” heisst, ermitteln wir dies mit der nachstehenden Codezeile:

```
langenscheidt["trump"]
```

```
'erdichten- schwindeln- sich aus den Fingern saugen'
```

Für R Nutzer

Eine Python *Dictionary* entspricht einer R *List*. Diese werden in R nicht gleich erstellt, dafür aber auf die gleiche Weise abgefragt wie *Dictionaries* in Python

5.1 Übung 1: Dictionary

Erstelle eine *Dictionary* mit folgenden Einträgen: Vorname und Nachname von (d)einer Person. Weise diese Dictionary der Variable `me` zu.

```
# Musterlösung  
  
me = {"vorname": "Guido", "nachname": "van Rossum"}
```

5.2 Übung 2: Elemente aus Dictionary ansprechen

Rufe verschiedene Elemente aus der Dictionary via dem *Key* ab.

```
me["vorname"]
```

```
'Guido'
```

```
me["nachname"]
```

```
'van Rossum'
```

5.3 Übung 3: Key ergänzen

Um einer *Dictionary* mit einem weiteren Eintrag zu ergänzen, geht man sehr ähnlich vor wie beim Abrufen von Einträgen.

```
langenscheidt["trumpet"] = "trompete"
```

Ergänze gemäss nachstehendem Beispiel die Variable `me` durch den Eintrag `groesse`.

Für R Nutzer

Python *Dictionaries* werden nicht nur gleich abgerufen wie R *Lists*, sonder auch gleich ergänzt.

```
me["groesse"] = 181
```

5.4 Input

Ein *Key* kann auch mehrere Einträge enthalten. An unserem Langenscheidts Beispiel: Das Wort “trump” ist zwar eindeutig, doch “trumpery” hat vier verschiedene Bedeutungen. In so einem Fall können wir einem Eintrag auch eine *List* von Werten zuweisen. Beachte die Eckigen Klammern und die Kommas, welche die Listeneinträge voneinander trennt.

```
langenscheidt["trumpery"] = ["Plunder- Ramsch- Schund", "Gewäsch- Quatsch", "Schund-  
↪ Kitsch", "billig- nichtssagend"]  
langenscheidt["trumpery"]
```

```
['Plunder- Ramsch- Schund',  
 'Gewäsch- Quatsch',  
 'Schund- Kitsch',  
 'billig- nichtssagend']
```

5.5 Übung 4: Dictionary mit List

Erstelle eine neue Dictionary mit den gleichen Keys wie *me*, aber diesmal mit mehreren Einträgen pro *Key* (also mehreren Vornamen, Nachnamen usw.). Beachte, dass nun jeder Eintrag eine *List* sein muss. Weise diese Dictionary der Variabel *people* zu.

```
people = {"vornamen": ["Christopher", "Henning", "Severin"], "nachnamen": ["Annen",  
↪ "May", "Kantereit"], "groessen": [174, 182, 162]}
```

5.6 Übung 5: Functions

Bisher haben wir Objekte erstellt. Richtig interessant wird programmieren aber erst, wenn wir mit Objekte durch Funktionen (*Functions*) verändern. *Functions* führen bestimmte Tasks aus. Wende die Functions *len()* und *sum()* an den Listen *groessen* und *vornamen* an.

```
# Musterlösung
```

```
len(groessen)
```

```
-----  
NameError                                Traceback (most recent call last)  
<ipython-input-11-2853382417e4> in <module>  
      1 # Musterlösung  
      2  
----> 3 len(groessen)  
  
NameError: name 'groessen' is not defined
```

```
# Musterlösung
```

```
sum(groessen)
```

```
695
```

```
# Musterlösung
```

```
len(vornamen)
```

```
4
```

```
# Musterlösung
```

```
# sum(vornamen)
```

INPUT: PYTHON MODULES

Ähnlich wie R basiert Python auf Erweiterungen: Diese Erweiterungen heissen in R *Libraries* oder *Packages*, in Python werden sie *Modules* genannt. Sie sind dazu da, gewisse Teilbereiche unseres Arbeitsprozesses zu vereinfachen. Eine Analogie dazu: Um ein Haus zu bauen sind wir auf verschiedene Spezialisten / Spezialistinnen angewiesen: Wir brauchen zum Beispiel eine Malerin oder einen Maler. Im Telefonbuch sind seitenweise Maler*innen aufgelistet, und jede*r arbeitet etwas anders. Um eine spezifische Malerin anzuheuern müssen wir zuerst den Kontakt herstellen und die Vertragsmodalitäten vereinbaren. Erst dann können wir sie in unseren Arbeitsprozess (Haus bauen) einbinden. Um diese Analogie auf unser Projekt zu übertragen: Das “Haus bauen” ist unser Forschungsprojekt (z.B. eine Bachelorarbeit). Ein “Telefonbuch”, wo die Spezialisten erfasst sind nennt ein *Repository*.

6.1 Erweiterung installieren

Den Erstkontakt mit der Malerin zu erstellen und die Vertragsmodalitäten zu vereinbaren bedeutet, die Erweiterung zu installieren. In R ist die Installation einer *Library* ein R-Befehl und wird innerhalb von R ausgeführt. In Python ist dies leider etwas komplizierter, es braucht für die Installation einer Python library eine Zusatzsoftware. Diese lernen wir zu einem späteren Zeitpunkt kennen, schauen wir uns jetzt nur mal den R-Code zur Installation eines Packages an:

R	python
<code>install.packages("malerinWanger")</code>	<code># Module können nicht innerhalb # von Python installiert werden</code>

In diesem Beispiel heisst die Erweiterung `malerinWanger`. Das *Repository* geben wir in R meistens nicht an, weil in RStudio bereits eine Default-Adresse hinterlegt ist.

6.2 Erweiterung laden

Treiben wir an dieser Stelle die Analogie etwas weiter: Der Erstkontakt mit der Malerin ist also erstellt und alle Vertragsmodalitäten sind vereinbart. Nun wollen wir an einem bestimmten Tag mit ihr arbeiten. Dafür müssen wir sie zuerst auf die Baustelle bestellen. Übersetzt auf Programmieren bedeutet dies, wir müssen die Erweiterung in unsere Session laden. In R und Python sehen die Befehle folgendermassen aus:

R	python
<code>library(malerlinWanger)</code>	<code>import malerinWanger</code>

6.3 Erweiterung verwenden

Erst jetzt können wir mit der Erweiterung arbeiten und die Fachexpertise unserer Malerin nutzen. Eine Expertise unserer Malerin ist es, Wände zu bemalen. Dafür gibt es eine *Function* `wand_bemalen()`. In R kann ich diese *Function* “einfach so” aufrufen. In Python hingegen muss ich die Erweiterung, in der die *Function* enthalten ist, der *Function* mit einem Punkt voranstellen. Das sieht also folgendermassen aus:

R	python
<code>wand_bemalen()</code>	<code>malerinWanger.wand_bemalen()</code>

Das ist zwar umständlicher, aber dafür weniger Fehleranfällig. Angenommen, unser Maurer kann ebenfalls Wände bemalen und hat die entsprechende *Function* `wand_bemalen()` ebenfalls. Dann ist in R nicht klar, welche Erweiterung gemeint ist und das kann zu Missverständnissen führen (vielleicht bemalt der Maurer die Wände etwas anders als die Malerin). In Python ist im obigen Beispiel unmissverständlich, dass ich `wand_bemalen()` aus dem Modul `malerinWanger` meine.

6.4 Modul mit Alias importieren

Da es umständlich sein kann, jedesmal `malerinWanger.wand_bemalen()` voll auszuschreiben, können wir beim Importieren dem Modul auch einen “Alias” vergeben. Dies kann beispielsweise folgendermassen aussehen:

```
import malerinWanger as mm
mm.wand_bemalen()
```

Dies ist deshalb wichtig, weil sich für viele Module haben sich bestimmte Aliasse eingebürgert haben. Ihr macht sich das Leben leichter, wenn ihr euch an diese Konventionen (welche ihr noch kennenlernen werdet) hält.

6.5 Einzelne *Function* importieren

Es gibt noch die Variante, eine explizite *Function* aus einem Modul zu laden. Wenn man dies macht, kann man die Funktion ohne vorangestelltes Modul nutzen (genau wie in R). Dies sieht folgendermassen aus:

```
from malerinWanger import wand_bemalen
wand_bemalen()
```

6.6 Alle *Functions* importieren

Zusätzlich ist es möglich, **alle** *Functions* aus einem Modul zu importieren. Diese Notation wird nicht empfohlen und ist hier nur erwähnt, weil ihr allenfalls diese Schreibweise antreffen werdet.

```
from malerinWanger import *
wand_bemalen()
```

Das wichtigste in Kürze

- Erweiterungen heissen in Python *Modules*

- Vor der erstmaligen Nutzung muss ein *Module* installiert werden
 - Vor der Verwendung in einer Session muss ein *Module* importiert werden
 - Ein *Module* wird in Python mit `import modulename` geladen
 - Module müssen für **jede Session** importiert werden!
 - Eine *Function* aus einem *Module* wird folgendermassen aufgerufen: `modulname.function()`
 - Ein Modul kann auch mit einem Alias importiert werden `import modul as modulalias`
 - Eine Einzelne *Function* kann aus einem Modul explizit importiert werden: `from malerinWanger import wand_bemalen`
 - Mit einem Platzhaltersymbol `*` können auch alle *Functions* aus einem Modul importiert werden: `from malerinWanger import *`
-

ÜBUNG: TABELLARISCHE DATEN

Schauen wir uns nochmals die *Dictionary* `people` an. Diese ist ein Spezialfall einer *Dictionary*: Jeder Eintrag besteht aus einer Liste von gleich vielen Werten. Wie bereits erwähnt, kann es in einem solchen Fall sinnvoll sein, die *Dictionary* als Tabelle darzustellen. Dafür brauchen wir das Modul `pandas`, welches wir mit dem Alias `pd` importieren.

7.1 Übung 1: von einer *Dictionary* zu einer *DataFrame*

Wandle die *Dictionary* `people` in eine *DataFrame* um: Dazu musst du `people` als Argument der Funktion `DataFrame` übergeben: `pd.DataFrame(people)`. Weise den Output der Variable `people_df` zu.

Für R Nutzer: Auch in R gibt es *Dataframes*, diese sind aber in Base R integriert (brauchen deswegen keine Erweiterung wie `pandas`)

7.2 Übung 2: *DataFrame* in csv umwandeln

In der Praxis kommen Tabellarische Daten meist als «csv» Dateien daher. Wir können aus unserer eben erstellten *DataFrame* sehr einfach eine csv Datei erstellen. Führe das mit folgendem Code aus:

```
people_df.to_csv("people.csv")
```

7.3 Übung 3: CSV als *DataFrame* importieren

Genau so einfach ist es eine csv zu importieren. Lade dazu die Datei “zeckenstiche.csv” von Moodle herunter und speichere es im aktuellen Arbeitsverzeichnis ab. Importiere mit folgendem Code die Datei “zeckenstiche.csv”. Schau dir `zeckenstiche` nach dem importieren an.

```
zeckenstiche = pd.read_csv("zeckenstiche.csv")
```

Hinweis: Dieser Code funktioniert nur, wenn “zeckenstiche.csv” im aktuellen Arbeitsverzeichnis (*Current Working Directory*) abgespeichert wird. Wenn du nicht sicher bist, wo dein aktuelles Arbeitsverzeichnis liegt, kannst du das Modul `os` laden und dies mit der Funktion `os.getcwd()` (get *current working directory*) herausfinden.

```
import os  
  
os.getcwd()
```

```
'/home/runner/work/codingingis/codingingis'
```

7.4 Übung 4: Koordinaten räumlich darstellen

Die *DataFrame* `zeckenstiche` beinhaltet x und y Koordinaten für jeden Unfall in den gleichnamigen Spalten. Wir können die Stiche mit einem Scatterplot räumlich visualisieren. Führe dazu folgenden Code aus:

```
%matplotlib inline # Könnt ihr ignorieren

fig = zeckenstiche.plot.scatter("x", "y")

fig.axis("scaled")
```

```
UsageError: unrecognized arguments: # Könnt ihr ignorieren
```

7.5 Übung 5: Einzelne Spalte selektieren

Um eine einzelne Spalte zu selektieren (z.B. die Spalte "ID") kann man gleich vorgehen wie bei der Selektion eines Eintrags in einer *Dictionary*. Probiere es aus.

```
zeckenstiche["ID"]
```

```
0          0
1          1
2          2
3          3
4          4
...
1071      1071
1072      1072
1073      1073
1074      1074
1075      1075
Name: ID, Length: 1076, dtype: int64
```

7.6 Übung 6: Neue Spalte erstellen

Auch das Erstellen einer neuen Spalte ist identisch mit der Erstellung eines neuen *Dictionary* Eintrags. Folgende Zeile erstellt eine neue Spalte "Stichtyp" mit dem gleichen Wert in allen Zeilen ("Zecke"). Erstelle ebenfalls eine solche Spalte.

```
zeckenstiche["Stichtyp"] = "Zecke"

zeckenstiche
```

	ID	accuracy	x	y	Stichtyp
0	0	65.617154	2678971.0	1240824.0	Zecke
1	1	257.492052	2679837.0	1240858.0	Zecke

(continues on next page)

(continued from previous page)

2	2	163.533834	2687539.0	1240881.0	Zecke
3	3	185.000000	2674848.0	1240913.0	Zecke
4	4	228.215231	2681771.0	1240922.0	Zecke
...
1071	1071	109.531946	2678005.0	1257344.0	Zecke
1072	1072	100.489274	2678005.0	1257347.0	Zecke
1073	1073	301.748529	2689893.0	1257351.0	Zecke
1074	1074	301.748542	2690668.0	1257369.0	Zecke
1075	1075	226.000000	2682852.0	1257531.0	Zecke

[1076 rows x 5 columns]

EINLEITUNG ZU DIESEM BLOCK

Letzte Woche haben wir einen weiten Bogen gespannt: Von einfachen (primitiven) Datentypen bis hin zu tabellarischen Daten. In dieser und der nächsten Blöcken (Coding in GIS II und III) wollen wir uns weiter mit den Zeckenstichdaten befassen. Wir werden im Wesentlichen ein Teil der Übung aus “Datenqualität und Unsicherheit” in Python rekonstruieren.

In der Übung geht es um folgendes: Wir wissen das die Lagegenauigkeit der Zeckenstichmeldungen mit einer gewissen Unsicherheit behaftet sind. Um die Frage “Welcher Anteil der der Zeckenstiche befinden sich im Wald?” unter Berücksichtigung dieser Unsicherheit beantworten zu können, führen wir eine (Monte Carlo) Simulation durch. In dieser Simulation verändern wir die Position der Zeckenstichmeldungen zufällig und berechnen den Anteil der Zeckenstiche im Wald. Das zufällige Verschieben und berechnen wiederholen wir beliebig lange und bekommen für jede Wiederholung einen leicht unterschiedlichen Prozentwert. Die Verteilung dieser Prozentwerte ist die Antwort auf die ursprüngliche Frage (“Welcher Anteil...”) unter Berücksichtigung der Unsicherheit.

Um eine solche, etwas komplexere Aufgabe lösen zu können müssen wir sie in einfachere Einzelschritte aufteilen. Diese bearbeiten wir in dieser und der kommenden Woche:

- Schritt 1: Einen Einzelpunkt zufällig verschieben
- Schritt 2: Alle Punkte einer DataFrame zufällig verschieben (1 “Run”)
- Schritt 3: Alle Punkte einer DataFrame mehrfach zufällig verschieben (z.B. 50 “Runs”)
- Schritt 4: Anteil der Punkte im Wald pro “Run” ermitteln
- Schritt 5: Verteilung dieser Mittelwerte visualisieren

Übungsziele

- Functions kennenlernen und beherrschen
 - Einfache Geometrien manipulieren lernen
 - Eigene Python-Scripts importieren können
 - Function auf eine ganze Spalte einer (Geo-) DataFrame anwenden können.
-

INPUT: *FUNCTION* BASICS

Bevor wir uns in die Simulation stürzen müssen wir noch lernen eigene *Functions* zu schreiben: Dies ist auch nicht weiter schwierig: Eine *Function* wird mit `def` eingeleitet, braucht einen Namen, einen Input und einen Output.

Wenn wir zum Beispiel eine Function erstellen wollen die uns grüßt, so geht dies folgendermassen:

```
def sag_hallo():  
    return "Hallo!"
```

- Mit `def` sagen wir: “Jetzt definiere ich eine Function”.
- Danach kommt der Name der *Function*, in unserem Fall `sag_hallo` (mit diesem Namen können wir die *Function* später wieder abrufen).
- Als drittes kommen die runden Klammern, wo wir bei Bedarf Inputvariablen (sogenannte Parameter) festlegen können. In diesem ersten Beispiel habe ich keine Parameter festgelegt
- Nach der Klammer kommt ein Doppelpunkt was bedeutet: “jetzt wird gleich definiert, was die Funktion tun soll”
- Auf einer neuen Zeile wird eingerückt festgelegt, was die Function eben tun soll. Meist sind hier ein paar Zeilen Code vorhanden
- Die letzte eingerückte Zeile (in unserem Fall ist das die einzige Zeile) gibt mit `return` an, was die *Function* zurück geben soll (der Output). In unserem Fall soll sie “Hallo!” zurück geben.
- Voila, das war’s schon! Jetzt können wir diese Function schon nutzen:

```
sag_hallo()
```

```
'Hallo!'
```

Diese *Function* ohne Input ist wenig nützlich. Meist wollen wir der *Function* etwas - einen Input - übergeben können. Um eine *Function* zu erstellen die ein Argument annimmt geht man folgendermassen vor:

```
def sag_hallo(vorname):  
    return "Hallo "+vorname+"!"
```

Nun können wir der Function einen Parameter übergeben, damit wir persönlich begrüßt werden. In folgendem Beispiel ist `vorname` ein Parameter, “Guido” ist sein Argument.

```
sag_hallo("Guido")
```

```
'Hallo Guido!'
```

Den Output können wir wie gewohnt einer neuen Variabel zuweisen:

```
persoenlicher_gruss = sag_hallo("Guido")  
persoenlicher_gruss
```

```
'Hallo Guido!'
```


ÜBUNG: *FUNCTIONS*

10.1 Übung 1: Erste *Function* erstellen

Erstelle eine *Function*, die `gruezi` heisst, einen Nachnamen als Input annimmt und per Sie grüsst. Das Resultat soll in etwa folgendermassen aussehen:

```
# Musterlösung

def gruezi(nachname):
    return("Guten Tag, "+nachname)
```

```
gruezi("Guido")
```

```
'Guten Tag, Guido'
```

10.2 Übung 2: *Function* erweitern

Erweitere die *Function* `gruezi` indem eine du einen weiteren Parameter namens `anrede` implementierst. Das Resultat soll in etwa folgendermassen aussehen:

```
# Musterlösung

def gruezi(nachname, anrede):
    return("Guten Tag, "+anrede+" "+nachname)
```

```
gruezi("van Rossum", "Herr")
```

```
'Guten Tag, Herr van Rossum'
```

10.3 Übung 3: Default-Werte festlegen

Man kann für die Parameter folgendermassen einen Standardwert festlegen: Beim Definieren der Function wird dem Parameter schon innerhalb der Klammer ein Argument zugewiesen (z.B. `anrede = "Herr oder Frau"`). Wenn `anrede` bei der Verwendung von `gruezi` nicht definiert wird, entspricht die Anrede nun «Herr oder Frau». Setze einen Standardwert in der Anrede und teste die *Function*. Das Resultat soll in etwa folgendermassen aussehen:

```
# Musterlösung
```

```
def gruezi(nachname, anrede = "Herr oder Frau"):
    return("Guten Tag, "+anrede+" "+nachname)
```

```
gruezi("van Rossum")
```

```
'Guten Tag, Herr oder Frau van Rossum'
```

INPUT: *FUNCTION* ADVANCED

11.1 *Functions* mit Default Werten

Wenn Parameter Default-Argumente zugewiesen werden, dann werden sie für den Nutzer automatisch zu optionalen Parametern. Wenn eine neue *Function* erstellt wird, kommen die Optionalen Parameter immer am Schluss. Also so:

```
def gruezi(nachname, anrede = "Herr oder Frau"):
    return("Guten Tag, "+anrede+" "+nachname)
```

Wenn wir versuchen die optionalen Parameter an den Anfang zu stellen, meldet sich Python mit einem Error:

```
def gruezi(anrede = "Herr oder Frau", nachname):
    return("Guten Tag, "+anrede+" "+nachname)
```

Error:

```
File "<ipython-input-12-b48869b8c585>", line 1
    def gruezi(anrede = "Herr oder Frau", nachname):
        ^
SyntaxError: non-default argument follows default argument
```

11.2 Reihenfolge der Argumente

Wenn die richtige Reihenfolge eingehalten wird, müssen die Parameter (z.B: anrede=, nachname=) nicht spezifiziert werden. Zum Beispiel:

```
gruezi("van Rossum", "Herr")
```

```
'Guten Tag, Herr van Rossum'
```

Wenn wir die Reihenfolge verändern, ist der Output unserer Funktion fehlerhaft:

```
gruezi("Herr", "van Rossum")
```

```
'Guten Tag, van Rossum Herr'
```

Aber wenn die Parameter der Argumente spezifiziert werden, ist die Reihenfolge wiederum egal:

```
gruezi(anrede = "Herr", nachname = "van Rossum")
```

```
'Guten Tag, Herr van Rossum'
```

11.3 Globale und Lokale Variablen

Innerhalb einer *Function* können nur die Variablen verwendet werden, die der *Function* als Argumente übergeben werden (oder innerhalb der *Funktion* erstellt werden). Diese nennt man “lokale” Variablen, sie sind lokal in der *Function* vorhanden. Im Gegensatz dazu stehen “globale” Variablen, diese sind teil der aktuellen Session. Folgende Vorgehensweise erzeugt also eine Fehlermeldung:

```
# Auch wenn wir die Variablen, welche gruezi() in der Funktion braucht  
# als globale Variablen definieren erhalten wir eine Fehlermeldung (s.u.)  
nachname = "van Rossum"  
anrede = "Herr"  
  
gruezi()
```

Error:

```
-----  
TypeError                                Traceback (most recent call last)  
<ipython-input-16-fc94d019a607> in <module>()  
      2 anrede = "Herr"  
      3  
----> 4 gruezi()  
  
TypeError: gruezi() missing 1 required positional argument: 'nachname'
```

ÜBUNG: EINEN EINZELPUNKT ZUFÄLLIG VERSCHIEBEN

Um unsere Monte Carlo Simulation umzusetzen, müssen wir die erwähnten Schritte 1-5 umsetzen. In diesem Teil packen Schritt 1 an, “Einen Einzelpunkt zufällig verschieben”.

- **Schritt 1: Einen Einzelpunkt zufällig verschieben**
- Schritt 2: Alle Punkte einer DataFrame zufällig verschieben (1 “Run”)
- Schritt 3: Alle Punkte einer DataFrame mehrfach zufällig verschieben (z.B. 50 “Runs”)
- Schritt 4: Anteil der Punkte im Wald pro “Run” ermitteln
- Schritt 5: Verteilung dieser Mittelwerte visualisieren

Dazu schauen wir uns die Zeckenstich-Tabelle nochmal an:

```
zeckenstiche
```

	ID	accuracy	x	y
0	0	65.617154	2678971.0	1240824.0
1	1	257.492052	2679837.0	1240858.0
2	2	163.533834	2687539.0	1240881.0
3	3	185.000000	2674848.0	1240913.0
4	4	228.215231	2681771.0	1240922.0

Wir haben also x und y -Werte, sowie eine “accuracy” Angabe. Um unsere Methode zu entwickeln, erstellen wir am besten einen “Dummy”-Punkt. Beispielsweise können wir den Testpunkt mit den Koordinaten der alten Sternwarte Bern (bzw. deren Gedenktafel, s.u.) erstellen.

```
x = 2600000  
y = 1200000  
accuracy = 60
```

Dieser konstruierte Punkt liegt im neuen Schweizer Koordinatensystem (CH1903+ LV95) auf der alten Sternwarte in Bern, bzw. auf deren Gedenktafel.



Fig. 12.1: Links: Koordinatensystem der Schweiz, mit dem (alten) Referenzwert Sternwarte Bern. Quelle: lv95.bve.be.ch rechts: Gedenktafel an die Alte Sternwarte Bern. Quelle: aiub.unibe.ch

12.1 Übung 1: Zufallswerte generieren

Es sind verschiedene Methoden denkbar, wie wir ein neues x,y Koordinatenpaar aus den bestehenden Koordinaten bestimmen. Die einfachste Variante ist es wohl, zu jedem Achsenwert (x/y) einen Zufallswert zu addieren. Sagen wir mal im Bereich von ± 100 m.

Wie generiert man aber Zufallszahlen in Python? Aus welcher Verteilung sollten die Zufallswerte stammen?

```
# Musterlösung

import random

# Uniforme verteilte Werte (zw. -100 bis +100)
random.randrange(-100,100)

# Normalverteilte Werte mit
# Mittelwert 0 und
# Standardabweichung 100
# Achtung: bei dieser Standardabweichung sind ca 30% der Werte > 100!
random.normalvariate(0,100)
```

8.532319327362256

12.2 Übung 2: Zufallswerte addieren

Addiere nun die Zufallszahlen zu den Dummy-Werten um `x_neu` und `y_neu` zu erhalten.

```
# Musterlösung
```

```
x_neu = x+random.normalvariate(0,100)
y_neu = y+random.normalvariate(0,100)
```

```
print(x_neu, y_neu)
```

```
2600094.1660849387 1199925.661278094
```

12.3 Übung 3: Arbeitsschritte in eine *Function* verwandeln

Jetzt sind die Einzelschritte zur Verschiebung eines Punktes klar. Da wir dies für viele Punkte machen müssen, ist es sinnvoll, aus den Arbeitsschritten eine *Function* zu erstellen. Erstelle eine *Function* `point_offset` welche als Input eine `x` oder `y` Koordinate annimmt und eine leicht verschobene Koordinate zurück gibt. Wenn du möchtest kannst du die Distanz der Verschiebung als optionalen Parameter (mit *Default* = 100) definieren.

```
# Musterlösung
```

```
def offset_coordinate(old, distance = 100):
    new = old + random.normalvariate(0,distance)

    return (new)
```

```
offset_coordinate(x, 100)
```

```
2599946.9286137717
```

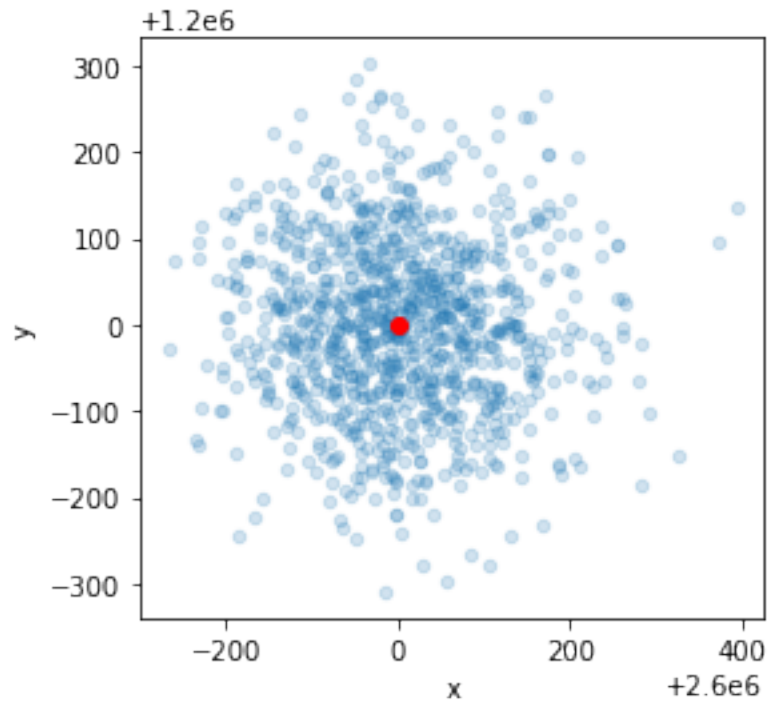
12.4 Übung 4: Output visualisieren

Nun ist es wichtig, dass wir unser Resultat visuell überprüfen. Im Beispiel unten wende ich die *Function* `offset_coordinate()` 1000x auf die Koordinaten der alten Sternwarte an. Für diesen Schritt gebe ich euch den fertigen Code, da ihr die dafür benötigten Techniken noch nicht gelernt habt. Füge diesen Code in dein Script ein und führe ihn aus. Allenfalls musst du den Code leicht an deine Situation anpassen.

```
from matplotlib import pyplot as plt
```

```
fig = pd.DataFrame({"x": [offset_coordinate(x) for rand in range(1,1000)], "y": [
↳ offset_coordinate(y) for rand in range(1,1000)]}).plot.scatter("x", "y", alpha = 0.
↳ 2)
```

```
plt.scatter(x,y, c = "red")
fig.axis("scaled")
plt.show()
```



ÜBUNG PUNKTE EINER *DATAFRAME* ZUFÄLLIG VERSCHIEBEN

Jetzt wo wir wissen, wie man einen einzelnen Punkt zufällig verschiebt können wir unsere *Function* auf die richtigen Daten, die Zeckenstiche, anwenden. Die ist Schritt 2 in unserem Schema:

- Schritt 1: Einen Einzelpunkt zufällig verschieben ✓
- **Schritt 2: Alle Punkte einer DataFrame zufällig verschieben (1 “Run”)**
- Schritt 3: Alle Punkte einer DataFrame mehrfach zufällig verschieben (z.B. 50 “Runs”)
- Schritt 4: Anteil der Punkte im Wald pro “Run” ermitteln
- Schritt 5: Verteilung dieser Mittelwerte visualisieren

Lade hierzu die nötigen Module und importiere die Zeckenstichdaten

```
import pandas as pd
import random

def offset_coordinate(old, distance = 100):
    new = old + random.normalvariate(0,distance)

    return (new)

# mit "head(10)" importiere ich nur die ersten 10 Einträge des CSV. So kann ich den
↪Workflow erstmal an einem kleineren Datensatz testen.
zeckenstiche = pd.read_csv("zeckenstiche.csv").head(10)

zeckenstiche
```

	ID	accuracy	x	y
0	0	65.617154	2678971.0	1240824.0
1	1	257.492052	2679837.0	1240858.0
2	2	163.533834	2687539.0	1240881.0
3	3	185.000000	2674848.0	1240913.0
4	4	228.215231	2681771.0	1240922.0
5	5	257.492053	2690119.0	1240932.0
6	6	17.951638	2678477.0	1240956.0
7	7	287.211395	2687505.0	1241029.0
8	8	301.748542	2681288.0	1241081.0
9	9	301.748542	2681268.0	1241111.0

13.1 Übung 1: Alle Zeckenstiche zufällig verschieben

Um in *DataFrames* / *GeoDataFrames* ganze Spalten oder Zeilen zu verändern, stehen einem die *Methods* `map`, `apply` und `applymap` zur Verfügung. Es lohnt sich sehr, die verschiedenen Anwendungsbereiche dieser *Methods* kennen zu lernen, doch das würde den Umfang dieses Kurses überschreiten. Heute brauchen wir lediglich `apply`, welches wir einmal auf die `x`- und einmal auf die `y`-Spalte unserer *DataFrame* anwenden. Um unsere eigene Funktion `offset_coordinate` auf die Spalten anzuwenden gehst du wie folgt vor: Selektiere die entsprechende Spalte (mit `["spaltenname"]`) und verwende die Method `apply` mit der selbst erstellten Funktion `point_offset` als einziges Argument, und zwar ohne Klammer.

```
# Musterlösung
```

```
zeckenstiche["x"].apply(offset_coordinate)
zeckenstiche["y"].apply(offset_coordinate)
```

```
0    1.240936e+06
1    1.240802e+06
2    1.240812e+06
3    1.240744e+06
4    1.240842e+06
Name: y, dtype: float64
```

13.2 Übung 2: Neue GeoDataFrame mit simulierten Punkten erstellen

Was bei `apply()` rauskommt, sind nur die neuen, verschobenen Punkte. Um diese abzuspeichern erstellen wir neue Spalten in `zeckenstiche` namen `x_sim`/`y_sim`:

```
# Musterlösung
```

```
zeckenstiche["x_sim"] = zeckenstiche["x"].apply(offset_coordinate)
zeckenstiche["y_sim"] = zeckenstiche["y"].apply(offset_coordinate)
```

```
zeckenstiche
```

	ID	accuracy	x	y	x_sim	y_sim
0	0	65.617154	2678971.0	1240824.0	2.678872e+06	1.241000e+06
1	1	257.492052	2679837.0	1240858.0	2.679613e+06	1.240965e+06
2	2	163.533834	2687539.0	1240881.0	2.687497e+06	1.241049e+06
3	3	185.000000	2674848.0	1240913.0	2.674818e+06	1.240775e+06
4	4	228.215231	2681771.0	1240922.0	2.681765e+06	1.240570e+06
5	5	257.492053	2690119.0	1240932.0	2.690078e+06	1.240843e+06
6	6	17.951638	2678477.0	1240956.0	2.678440e+06	1.241082e+06
7	7	287.211395	2687505.0	1241029.0	2.687479e+06	1.241010e+06
8	8	301.748542	2681288.0	1241081.0	2.681338e+06	1.241135e+06
9	9	301.748542	2681268.0	1241111.0	2.681112e+06	1.241050e+06

13.3 Übung 4: Mehrere *DataFrames* visualisieren

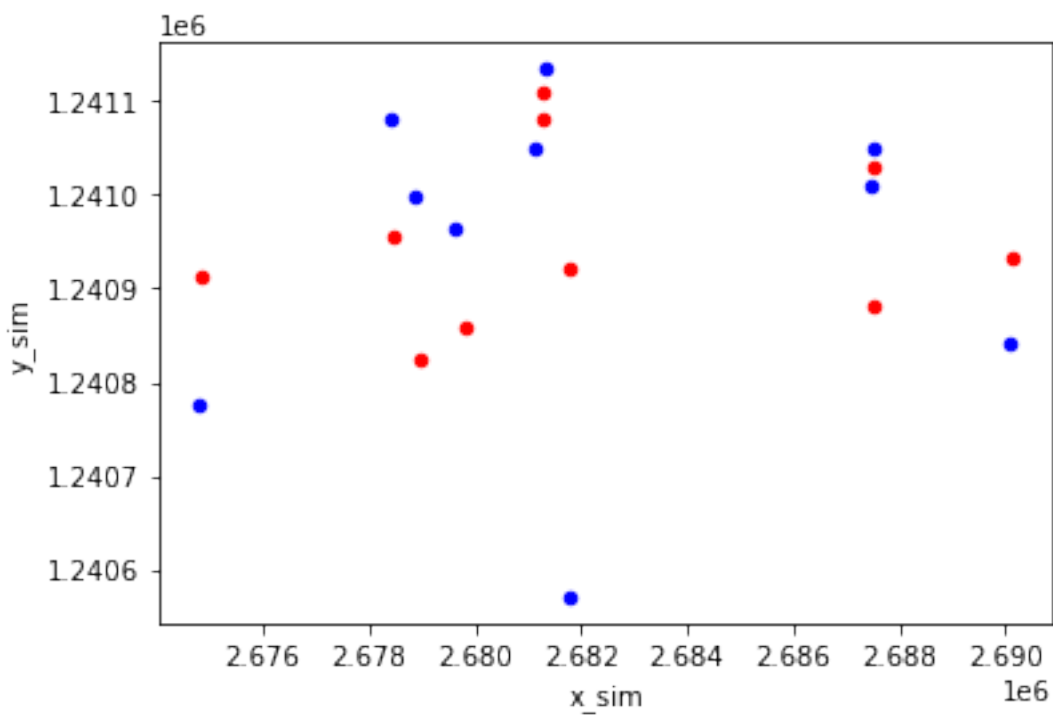
Um zwei *DataFrames* im gleichen Plot darzustellen, wird folgendermassen vorgegangen. Der erste Datensatz wird mit `.plot()` visualisiert, wobei der Output einer Variabel (z.B. `basemap`) zugewiesen wird. Danach wird der zweite Datensatz ebenfalls mit `.plot()` visualisiert, wobei auf den ersten Plot via dem Argument `ax` verwiesen wird.

```
%matplotlib inline # Könnt ihr ignorieren
```

```
# Musterlösung
```

```
basemap = zeckenstiche.plot.scatter("x", "y", color = "red")
zeckenstiche.plot.scatter("x_sim", "y_sim", ax = basemap, color = "blue")
```

```
<AxesSubplot:xlabel='x_sim', ylabel='y_sim'>
```



EINLEITUNG ZU DIESEM BLOCK

Übungsziele

- Ihr kennt For-Loops und könnt sie anwenden
 - Ihr verwendet eure erste räumliche Operation «Spatial Join» und wisst, dass es hier eine ganze Palette an weiteren Operatoren gibt
 - Ihr könnt eine (Geo-) DataFrame nach Gruppe Zusammenfassen
 - Ihr lernt weitere Visualisierungstechniken kennen
-

ÜBUNG: FOR LOOPS (TEIL I)

Nirgends ist der Aspekt der Automatisierung so sichtbar wie in For Loops. Loops sind «Schleifen» wo eine Aufgabe so lange wiederholt wird, bis ein Ende erreicht worden ist. Auch For-Loops sind im Grunde genommen sehr einfach:

```
for platzhalter in [0,1,2]:  
    print("Iteration",platzhalter)
```

```
Iteration 0  
Iteration 1  
Iteration 2
```

- `for` legt fest, dass eine For-Loop beginnt
- Nach `for` kommt eine Platzhalter-Variabel, die ihr beliebig benennen könnt. Im obigen Beispiel lautet diese `platzhalter`
- Nach dem Platzhalter kommt der Begriff `in`
- Nach `in` wird der “Iterator” festgelegt, also worüber der For-Loop iterieren soll (hier: über eine `List` mit den Werten `[0, 1, 2]`).
- Danach kommt ein Doppelpunkt : der Zeigt: “Nun legen wir gleich fest was im For-Loop passieren soll” (dies erinnert an die Konstruktion einer *Function*)
- Auf einer neuen Zeile wird eingerückt festgelegt, was in der For-Loop passieren soll. In unserem Fall wird etwas Nonsense in die Konsole ausgespuckt. Achtung: `return()` gibt’s in For-Loops nicht.

15.1 Übung 1: Erste For-Loop erstellen

Erstelle eine For-Loop, die über eine Liste von 3 Namen iteriert, und jede Person in der Liste grüßt (Output in die Konsole mittels `print`).

```
# Musterlösung  
  
for name in ["Il Buono", "Il Brutto", "Il Cattivo"]:  
    print("Ciao ",name)
```

```
Ciao  Il Buono  
Ciao  Il Brutto  
Ciao  Il Cattivo
```

15.2 Übung 2: For-Loop mit `range()`

Im Beispiel der Einführung iterieren wir über eine *List* mit den Werten `[0, 1, 2]`. Wenn wir aber über viele Werte iterieren wollen, ist es zu mühsam händisch eine Liste mit allen Werten zu erstellen. Mit `range(n)` erstellt Python ein Iterator mit den Zahlen 0 bis n. Repliziere den For-Loop aus der Einführung und ersetze `[0, 1, 2]` mit `range(3)`.

```
# Musterlösung

for platzhalter in range(3):
    print("Iteration",platzhalter)
```

```
Iteration 0
Iteration 1
Iteration 2
```


ÜBUNG: FOR LOOPS (TEIL II)

Bis jetzt haben wir lediglich Sachen in die Konsole herausgeben lassen, doch wie schon bei Functions ist der Zweck einer For-Loop meist, dass nach Durchführung etwas davon zurückbleibt. `return()` gibt es wie bereits erwähnt bei *For-Loops* nicht. Nehmen wir folgendes Beispiel:

```
for schlagwort in ["bitch", "lover", "child", "mother", "sinner", "saint"]:
    liedzeile = "I'm a " + schlagwort
    print(liedzeile)
```

```
I'm a bitch
I'm a lover
I'm a child
I'm a mother
I'm a sinner
I'm a saint
```

Der Output von dieser For-Loop sind zwar sechs Liederzeilen, wenn wir die Variabel `liedzeile` anschauen ist dort nur das Resultat aus der letzten Durchführung gespeichert. Das gleiche gilt auch für die variabel `schlagwort`.

```
liedzeile
```

```
"I'm a saint"
```

```
schlagwort
```

```
'saint'
```

Das verrät uns etwas über die Funktionsweise des *For-Loops*: Bei jedem Durchgang werden die Variablen immer wieder überschrieben. Wenn wir also den Output des ganzen For-Loops abspeichern wollen, müssen wir dies etwas vorbereiten. Dafür erstellen wir unmittelbar vor dem For-Loop einen leeren Behälter, zum Beispiel eine leere Liste (`strophe = []`). Nun können wir innerhalb des *Loops* `append()` nutzen, um den Output von einem Durchgang dieser Liste hinzu zu fügen.

```
strophe = []

for schlagwort in ["bitch", "lover", "child", "mother", "sinner", "saint"]:
    liedzeile = "I'm a " + schlagwort
    strophe.append(liedzeile)

strophe
```

```
["I'm a bitch",  
 "I'm a lover",  
 "I'm a child",  
 "I'm a mother",  
 "I'm a sinner",  
 "I'm a saint"]
```

16.1 Übung 1: Output aus For-Loop speichern

Erweitere deinen *For-Loop* aus Übung 1 so, dass der Output aller Durchgänge in einer Liste gespeichert werden.

```
# Musterlösung  
  
mylist = []  
  
for name in ["Il Buono", "Il Brutto", "Il Cattivo"]:  
    mylist.append("Ciao "+name)  
  
mylist
```

```
['Ciao Il Buono', 'Ciao Il Brutto', 'Ciao Il Cattivo']
```

ZECKENSTICH SIMULATION MIT LOOP

Nun geht es weiter mit unserer Zeckenstich Monte-Carlo Simulation. Schritte 1 und 2 haben wir bereits erledigt. Nun packen wir Schritt 3 an:

- Schritt 1: Einen Einzelpunkt zufällig verschieben ✓
- Schritt 2: Alle Punkte einer DataFrame zufällig verschieben (1 “Run”) ✓
- **Schritt 3: Alle Punkte einer DataFrame mehrfach zufällig verschieben (z.B. 50 “Runs”)**
- Schritt 4: Anteil der Punkte im Wald pro “Run” ermitteln
- Schritt 5: Verteilung dieser Mittelwerte visualisieren

Ladet dafür die nötigen Module (pandas und random), holt euch die Funktion `offset_point()` und importiert den Datensatz `zeckenstiche.csv`.

```
import pandas as pd
import random

def offset_coordinate(old, distance = 1000):
    new = old + random.normalvariate(0,distance)

    return (new)

zeckenstiche = pd.read_csv("zeckenstiche.csv").head(10)
```

17.1 Übung 1: Mit For-Loop `zeckenstiche` mehrfach verschieben

Um euer Gedächtnis etwas aufzufrischen: Letzte Woche hatten wir mit `apply()` sowie unserer eigenen *Function* `offset_coordinate` alle Koordinaten einer *DataFrame* verschoben und die neuen Koordinaten als neue Spalten abgespeichert.

```
zeckenstiche["x_sim"] = zeckenstiche["x"].apply(offset_coordinate)
zeckenstiche["y_sim"] = zeckenstiche["y"].apply(offset_coordinate)
zeckenstiche
```

	ID	accuracy	x	y	x_sim	y_sim
0	0	65.617154	2678971.0	1240824.0	2.677895e+06	1.240571e+06
1	1	257.492052	2679837.0	1240858.0	2.679589e+06	1.239450e+06
2	2	163.533834	2687539.0	1240881.0	2.687803e+06	1.242318e+06
3	3	185.000000	2674848.0	1240913.0	2.675107e+06	1.240632e+06
4	4	228.215231	2681771.0	1240922.0	2.681610e+06	1.240731e+06
5	5	257.492053	2690119.0	1240932.0	2.690661e+06	1.240027e+06

(continues on next page)

(continued from previous page)

6	6	17.951638	2678477.0	1240956.0	2.678394e+06	1.241446e+06
7	7	287.211395	2687505.0	1241029.0	2.688195e+06	1.240972e+06
8	8	301.748542	2681288.0	1241081.0	2.683148e+06	1.241769e+06
9	9	301.748542	2681268.0	1241111.0	2.681239e+06	1.239957e+06

Kombiniere dies nun mit deinem Wissen über Loops, um die Punkte der *DataFrame* nicht einmal, sondern 50 mal zu verschieben. Dazu brauchst du vor dem Loop eine leere Liste (z.B. `monte_carlo = []`) damit du den Output aus jedem Loop mit `append()` abspeichern kannst. Erstelle auch eine neue Spalte `Run_Nr` mit der Nummer der Durchführung (die du vom Platzhalter erhältst).

Fig. 17.1: Quelle [twitter.com](#)

```
# Musterlösung

monte_carlo = []
for i in range(50):
    zeckenstiche["x_sim"] = zeckenstiche["x"].apply(offset_coordinate)
    zeckenstiche["y_sim"] = zeckenstiche["y"].apply(offset_coordinate)
    zeckenstiche["Run_Nr"] = i
    monte_carlo.append(zeckenstiche)
```

17.2 Übung 2: DataFrames aus Simulation zusammenführen

Schau dir die Outputs an.

- Mit `type()`:
 - Was für ein Datentyp ist `zeckenstiche_sim`?
 - Was für ein Datentyp ist `monte_carlo`?
- Mit `len()`:
 - Wie vielen Elemente hat `zeckenstiche_sim`?

– Wie viele Elemente hat `monte_carlo`?

Worauf ich hinaus will: `zeckenstiche_sim` ist eine *DataFrame* und `monte_carlo` ist eine Liste von *DataFrames*. Glücklicherweise kann man eine Liste von ähnlichen *GeoDataFrames* (ähnlich im Sinne von: gleiche Spaltennamen und -typen) mit der Funktion `concat()` aus `pandas` zu einer einzigen *DataFrame* zusammenführen. Führe die Funktion aus und speichere den Output als `monte_carlo_df`.

```
type(zeckenstiche)
```

```
pandas.core.frame.DataFrame
```

```
type(monte_carlo)
```

```
list
```

```
len(zeckenstiche)
```

```
10
```

```
len(monte_carlo)
```

```
50
```

```
monte_carlo_df = pd.concat(monte_carlo)
```

17.3 Übung 3: Simulierte Daten visualisieren

Exploriere nun `monte_carlo_df`? Was ist es für ein Datentyp? Was hat es für Spalten? Visualisiere den Datensatz räumlich mit `monte_carlo_df.plot.scatter()`.

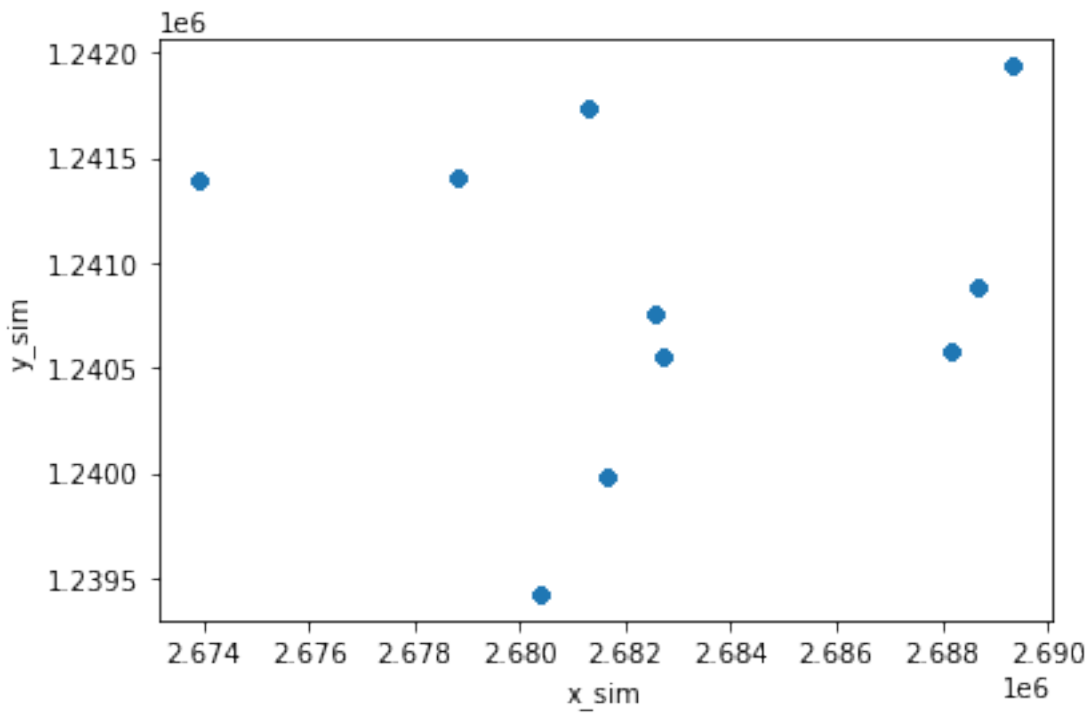
```
monte_carlo_df
```

	ID	accuracy	x	y	x_sim	y_sim	Run_Nr
0	0	65.617154	2678971.0	1240824.0	2.680389e+06	1.239422e+06	49
1	1	257.492052	2679837.0	1240858.0	2.681677e+06	1.239981e+06	49
2	2	163.533834	2687539.0	1240881.0	2.688693e+06	1.240889e+06	49
3	3	185.000000	2674848.0	1240913.0	2.673935e+06	1.241396e+06	49
4	4	228.215231	2681771.0	1240922.0	2.682692e+06	1.240552e+06	49
..
5	5	257.492053	2690119.0	1240932.0	2.689319e+06	1.241936e+06	49
6	6	17.951638	2678477.0	1240956.0	2.678810e+06	1.241407e+06	49
7	7	287.211395	2687505.0	1241029.0	2.688183e+06	1.240584e+06	49
8	8	301.748542	2681288.0	1241081.0	2.682548e+06	1.240764e+06	49
9	9	301.748542	2681268.0	1241111.0	2.681305e+06	1.241737e+06	49

```
[500 rows x 7 columns]
```

```
monte_carlo_df.plot.scatter("x_sim", "y_sim")
```

```
<AxesSubplot:xlabel='x_sim', ylabel='y_sim'>
```



ÜBUNG: GIS IN PYTHON

Die bisherigen Aufgaben hatten noch nicht viel mit GIS zu tun. Die Koordinaten der Zeckenstiche sind uns zwar bekannt, es handelt es dabei aber lediglich um Numerische xy-Werte, nicht um Geometrische Punkte mit einem Raumbezug. So bleiben uns bis jetzt alle räumlichen Operationen vorenthalten (zum Beispiel ob sich ein Stich im Wald befindet oder nicht), und wir können auch keine schönen Karten generieren.

Um mit Geodaten in Python arbeiten zu können, müssen wir ein neues Modul importieren. Im Grunde genommen sind Vektordaten nicht mehr als Tabellen mit einer zusätzlichen “Geometrie”-Spalte, dementsprechend baut die “Geo”-Erweiterung auf pandas auf und heisst: geopandas.

Du solltest geopandas bereits installiert haben und das Modul in dein Script importieren. Importiere geopandas mit dem Alias gpd.

18.1 Übung 1: *DataFrame* zu *GeoDataFrame*

Wie erwähnt sind die Zeckenstichdaten bisher lediglich als tabellarische Daten vorhanden. In ArcGIS Terminologie müssen wir die Operation “[XY Table to Point](#)” durchführen. In Python heisst das: Wir wandeln eine *DataFrame* in eine *GeoDataFrame* um. Zuerst erstellen wir eine Geometrie-Spalte aus den xy-Koordinaten mit der Funktion `points_from_xy` aus dem Modul geopandas.

```
zeckenstiche["geometry"] = gpd.points_from_xy(zeckenstiche.x, zeckenstiche.y)
```

Der Datensatz `zeckenstiche` hat jetzt aber noch nicht begriffen, dass es jetzt eine *GeoDataFrame* ist. Dies müssen wir dem Objekt erst noch mitteilen:

```
zeckenstiche = gpd.GeoDataFrame(zeckenstiche)
```

```
zeckenstiche
```

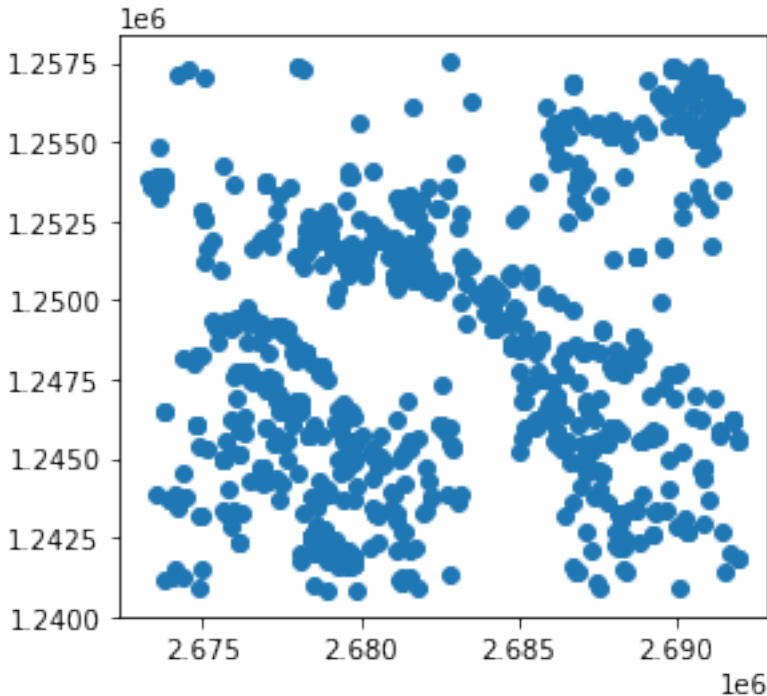
	ID	accuracy	x	y	geometry
0	0	65.617154	2678971.0	1240824.0	POINT (2678971.000 1240824.000)
1	1	257.492052	2679837.0	1240858.0	POINT (2679837.000 1240858.000)
2	2	163.533834	2687539.0	1240881.0	POINT (2687539.000 1240881.000)
3	3	185.000000	2674848.0	1240913.0	POINT (2674848.000 1240913.000)
4	4	228.215231	2681771.0	1240922.0	POINT (2681771.000 1240922.000)
...
1071	1071	109.531946	2678005.0	1257344.0	POINT (2678005.000 1257344.000)
1072	1072	100.489274	2678005.0	1257347.0	POINT (2678005.000 1257347.000)
1073	1073	301.748529	2689893.0	1257351.0	POINT (2689893.000 1257351.000)
1074	1074	301.748542	2690668.0	1257369.0	POINT (2690668.000 1257369.000)
1075	1075	226.000000	2682852.0	1257531.0	POINT (2682852.000 1257531.000)

```
[1076 rows x 5 columns]
```

Jetzt, wo zeckenstiche eine *GeoDataFrame* ist, gibt es einen einfachen Weg die Punkte räumlich zu visualisieren:

```
zeckenstiche.plot()
```

```
<AxesSubplot:>
```



18.2 Übung 2: Koordinatensystem festlegen

Wir wissen zwar, dass unsere *GeoDataFrame* in Schweizer Landeskoordinaten (CH1903 LV95) zu verstehen ist, aber dies haben wir noch nirgends festgehalten. Das Koordinatensystem (Coordinate Reference System, CRS) können wir über das Attribut `crs` der *GeoDataFrame* festhalten. Das Koordinatensystem CH1903 LV95 hat den EPSG Code 2056, demnach muss das CRS folgendermassen festgelegt werden:

```
zeckenstiche = zeckenstiche.set_crs(epsg = 2056)
```

Nun ist das Koordinatensystem (CRS) als Attribut der *GeoDataFrame* gespeichert:

```
zeckenstiche.crs
```

```
<Projected CRS: EPSG:2056>
Name: CH1903+ / LV95
Axis Info [cartesian]:
- E[east]: Easting (metre)
- N[north]: Northing (metre)
Area of Use:
- name: Europe - Liechtenstein and Switzerland
- bounds: (5.96, 45.82, 10.49, 47.81)
Coordinate Operation:
- name: Swiss Oblique Mercator 1995
```

(continues on next page)

(continued from previous page)

```
- method: Hotine Oblique Mercator (variant B)
Datum: CH1903+
- Ellipsoid: Bessel 1841
- Prime Meridian: Greenwich
```

18.3 Übung 3: Zeckenstiche als Shapefile exportieren

Zum Schluss exportieren wir unser Datensatz in ein Shapefile, damit wir das nächste Mal direkt mit einer *GeoDataFrame* arbeiten können. Genau wie wir in einer vorherigen Übung eine pandas DataFrame mit `.to_csv` in eine csv exportiert haben, gibt es für GeoDataFrames die Methode `.to_file`. Exportiere zeckenstiche mit dieser Methode in eine Shapefile.

```
zeckenstiche.to_file("zeckenstiche.shp")
```


WALDANTEIL BERECHNEN

Nun sind wir so weit, dass wir 50 Simulation der Zeckenstiche mit zufällig verschobenen Punkten vorbereitet haben. Wir haben also die gleiche Ausgangslage, mit der ihr den Themenblock «Datenqualität und Unsicherheiten» gestartet habt. Nun geht es darum:

1. Für jeden Simulierten Punkt zu bestimmen ob er innerhalb oder ausserhalb des Waldes liegt
2. Den Anteil der Punkte im Wald pro Simulation zu bestimmen
3. Die Verteilung dieser prozentualen Anteile über alle Simulationen zusammen zu fassen (mit einem Boxplot)

Lade dafür den Datensatz “Wald.zip” von Moodle herunter und entpacke das Shapefile in deine *Working Directory*. Importiere den Datensatz mit `pd.read_file()` und speichere es als Variable `wald`.

```
import pandas as pd
import geopandas as gpd

monte_carlo_df = gpd.read_file("monte_carlo_df.shp")

monte_carlo_df

wald = gpd.read_file("wald.shp")
zeckenstiche = gpd.read_file("zeckenstiche.shp")
```

19.1 Übung 1: Wald oder nicht Wald?

Als erstes stellt sich die Frage, welche Punkte sich innerhalb eines Wald-Polygons befinden. In GIS Terminologie handelt es sich hier um einen *Spatial Join*.

Spatial Join ist als Funktion im Modul `geopandas` mit dem namen `sjoin` vorhanden. Wie auf [der Hilfeseite](#) beschrieben, müssen wir dieser *Function* zwei *GeoDataFrames* übergeben, die ge-joined werden sollen. Es können weitere, optionale Parameter angepasst werden, doch bei uns passen die Default werte.

Führe `gpd.sjoin()` auf die beiden Datensätze `monte_carlo_df` und `wald` aus. Beachte, das die Reihenfolge, mit welchen du die beiden *GeoDataFrames* der Funktion übergibst eine Rolle spielt. Versuche beide Varianten und wähle die korrekte aus. Stelle dir dazu die Frage: Was für ein Geometrietyp (Punkt / Linie / Polygon) soll der Output haben? Speichere den Output als `monte_carlo_sjoin`. Hinweis: Allenfalls müssen das Koordinatensystem der beiden *GeoDataFrames* nochmals explizit gesetzt werden (z.B. mit `wald.set_crs(epsg = 2056, allow_override = True)`)

```
monte_carlo_df = monte_carlo_df.set_crs(epsg = 2056)
wald = wald.set_crs(epsg = 2056, allow_override = True)
```

(continues on next page)

(continued from previous page)

```
monte_carlo_sjoin = gpd.sjoin(wald,monte_carlo_df)

monte_carlo_sjoin.head()
```

```
-----
ImportError                                Traceback (most recent call last)
<ipython-input-2-b43b9d39760e> in <module>
      2 wald = wald.set_crs(epsg = 2056,allow_override = True)
      3
----> 4 monte_carlo_sjoin = gpd.sjoin(wald,monte_carlo_df)
      5
      6 monte_carlo_sjoin.head()

/opt/hostedtoolcache/Python/3.7.9/x64/lib/python3.7/site-packages/geopandas/tools/
↳ sjoin.py in sjoin(left_df, right_df, how, op, lsuffix, rsuffix)
      86         # keep them the same
      87         predicate = op
----> 88         sindex = right_df.sindex
      89         input_geoms = left_df.geometry
      90

/opt/hostedtoolcache/Python/3.7.9/x64/lib/python3.7/site-packages/geopandas/base.py_
↳ in sindex(self)
      591     def sindex(self):
      592         if not self._sindex_generated:
--> 593             self._generate_sindex()
      594         return self._sindex
      595

/opt/hostedtoolcache/Python/3.7.9/x64/lib/python3.7/site-packages/geopandas/base.py_
↳ in _generate_sindex(self)
      96
      97     def _generate_sindex(self):
----> 98         sindex_cls = get_sindex_class()
      99         if sindex_cls is not None:
     100             _sindex = sindex_cls(self.geometry)

/opt/hostedtoolcache/Python/3.7.9/x64/lib/python3.7/site-packages/geopandas/sindex.py_
↳ in get_sindex_class()
      41         return RTreeIndex
      42         raise ImportError(
----> 43             "Spatial indexes require either `rtree` or `pygeos`. "
      44             "See installation instructions at https://geopandas.org/install.html"
      45     )

ImportError: Spatial indexes require either `rtree` or `pygeos`. See installation_
↳ instructions at https://geopandas.org/install.html
```

19.2 Übung 2: Anteil der Punkte im Wald

Wenn wir mit `len()` die Anzahl Zeilen zwischen `monte_carlo_df` und `mote_carlo_sjoin` vergleichen, sehen wir, dass sich die Anzahl von 53'800 auf 24'021 reduziert hat. Im Mittel, über alle Simulationen, befinden sich also 44.5% ($\frac{24021}{53800}$) der Zeckenstiche im Wald. Das ist ja schon mal spannend, aber uns interessiert ja vor allem die Verteilung dieser Mittelwerte zwischen den Simulationen.

Wir müssen also die von `sjoin` übrig gebliebenen Punkte pro Durchgang zählen und durch ursprüngliche Anzahl Zeckenstiche dividieren. Ursprünglich waren es 1'076 Punkte, wie uns `len(zeckenstiche)` zeigt.

In der Logik: `DataFrame` => Nach «Durchgang» Gruppieren => zählen => dividieren. Genau in dieser Schreibweise können wir unsere Befehlskette aufbauen:

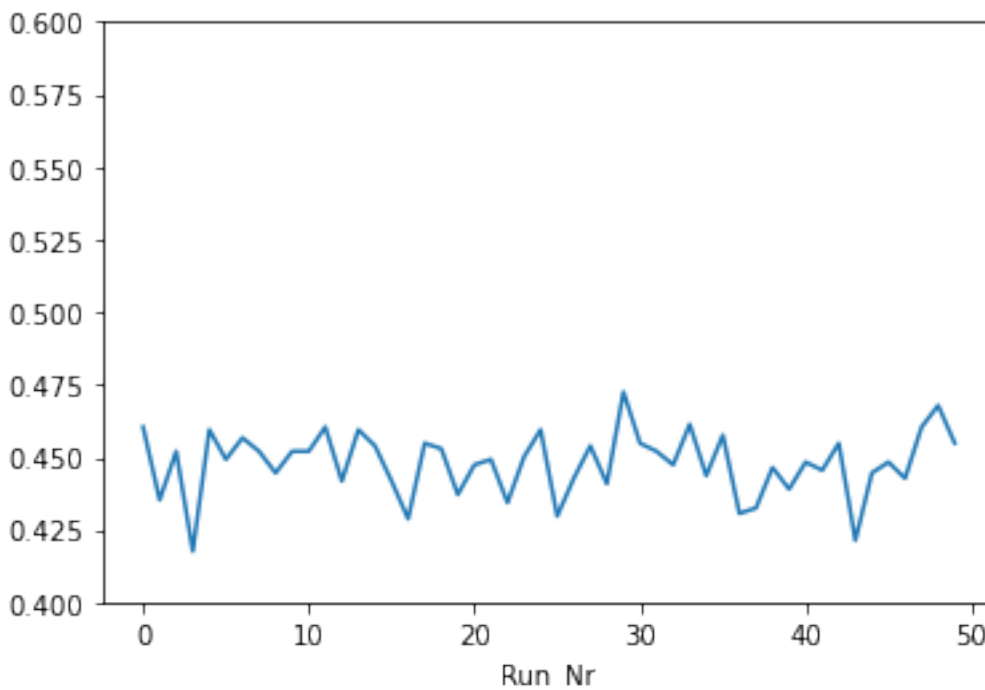
```
mittelwerte = monte_carlo_sjoin.groupby("Run_Nr").size()/1075
```

19.3 Übung 3: Mittelwerte Visualisieren

Gratuliere! Wenn du an diesem Punkt angekommen bist hast du eine ganze Monte Carlo Simulation von A bis Z mit Python durchgeführt. Von hier an steht dir der Weg frei für noch komplexere Analysen. Zum Abschluss kannst du die Mittelwerte wir nun auf einfache Weise visualisieren. Versuche dabei die Methods `plot()` und `plot.box()` sowie `plot.hist()`.

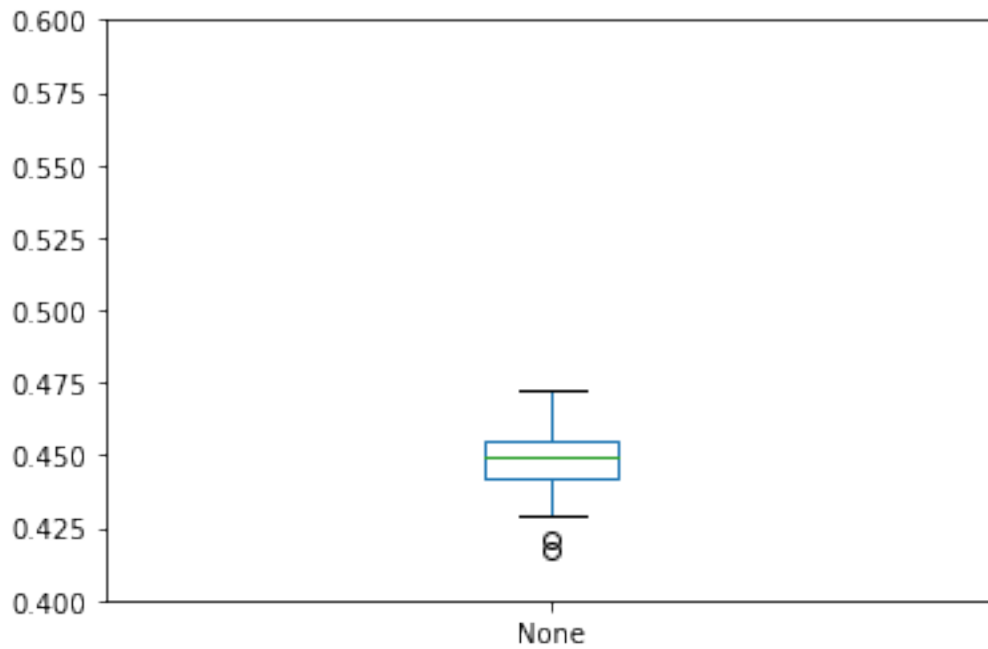
```
mittelwerte.plot().set_ylim(0.4,0.6)
```

```
(0.4, 0.6)
```



```
mittelwerte.plot.box().set_ylim(0.4,0.6)
```

```
(0.4, 0.6)
```



```
mittelwerte.plot.hist()
```

```
<AxesSubplot:ylabel='Frequency'>
```

