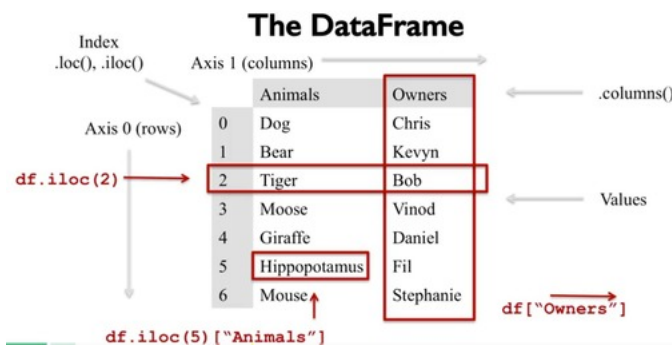


# 4. Pandas

## 1 PANDAS - IDEE

Datenanalyse Toolkit für Python

- designed um mit benannten (labeled) und voneinander abhängigen Daten zu arbeiten
- benutzt Numpy als Grundlage, hat dementsprechend Geschwindigkeitsvorteile gegenüber Python Built-In Methoden
- ermöglicht arbeiten mit Matrizen mit unterschiedlichen Datentypen (z.B. Strings und floats)
- matplotlib-Plot Funktionen sind eingebaut
- Zugriffsmöglichkeiten auf in der Matrix enthaltene Werte sehr ähnlich zu einem Python-Dictionary



- loc: labels (auch integer label), iloc: integer index

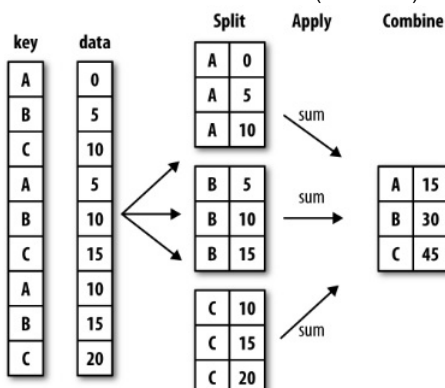
## 2 ELEMENTE VON PANDAS

- **Datenstrukturen** (hauptsächlich): 1D Array - Series, 2D Array - Dataframe
- **Index:** jede Achse hat nicht nur einen numerischen Index, sondern auch einen Namen. Indexe (Label) können auch in Multiindexen sortiert werden, die in hierarchischer Ordnung angeordnet sind.
- **groupby:** erlaubt einfaches und schnelles Zusammenfassen und Verarbeiten von zusammengehörigen Daten

## 3 VISUALISIERUNG VON DATAFRAME OPERATIONEN

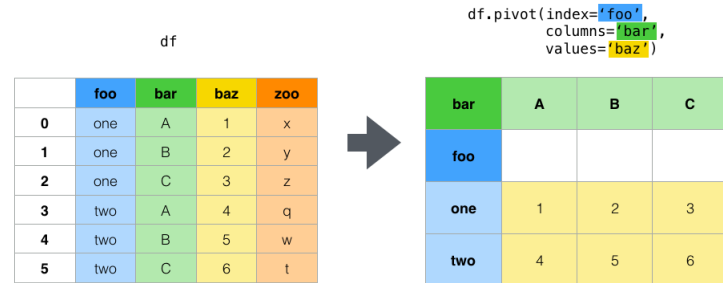
(aus der Pandas Dokumentation)

- **Groupby:** Zusammenfassen von Daten basierend auf Indexlabels (split) und Ausführen von Operationen (apply) von Daten, ein neuer Dataframe entsteht (combine)



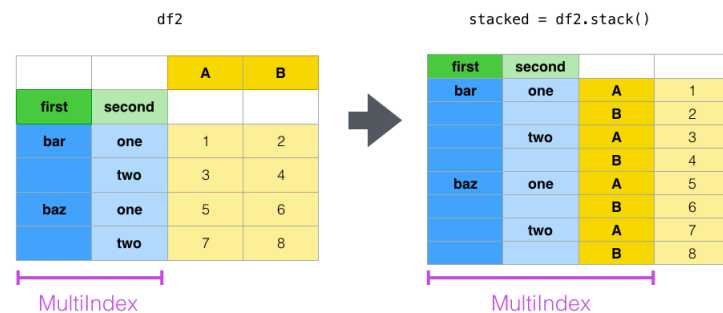
- **Pivot:** Verändern der Datenframe-Anordnung basierend auf den Labeln der Spalten (erstellen einer Pivot-Tabelle):

Pivot



- **Stack/Unstack:** Verändern der Struktur zwischen Reihenlabels und Spaltenlabels, Erzeugung von Multiindexen:

Stack



Unstack

