**PROBLEM SET 6**

Exercise 1. Data story

Ein Bild, das Screenshot enthält.

Automatisch generierte Beschreibung

Data Set: GDP per capita of Vietnam and China from 1986 to 2018

Ein Bild, das Text, Karte enthält.

Automatisch generierte Beschreibung

Story: By analysing correlation between 2 datasets, I figured out that Vietnam GDP per capita is having the most similarity to China GDP per capita with no time lag.

Exercise 2. Text mining with R

Text 1 : 20000-Leagues-Under-the-Sea.txt

Ein Bild, das Screenshot enthält.

Automatisch generierte Beschreibung

Text 2: The-War-of-the-Worlds.txt

Ein Bild, das Screenshot enthält.

Automatisch generierte Beschreibung

Ein Bild, das Text enthält.

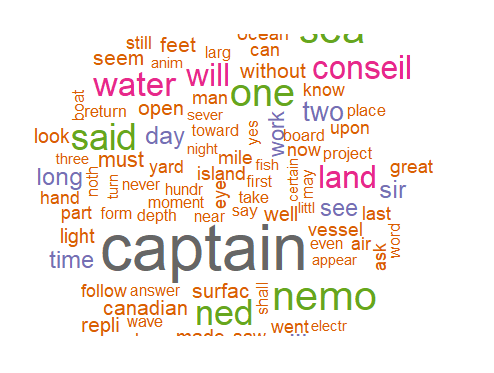
Automatisch generierte Beschreibung

Ein Bild, das Text, Zeitung enthält.

Automatisch generierte Beschreibung

Min frequency : 70

With color:



Exercise 3. Text mining with Python

Text 1: 20000-Leagues-Under-the-Sea.txt

Ein Bild, das Screenshot, Monitor, schwarz, Tisch enthält.

Automatisch generierte Beschreibung

Text 2: The-War-of-the-Worlds.txt

Ein Bild, das Screenshot, Monitor, schwarz, Computer enthält.

Automatisch generierte Beschreibung

Most Common Words

Ein Bild, das Screenshot, schwarz, Tisch, Monitor enthält.

Automatisch generierte Beschreibung

Exercise 4. Text mining with RapidMiner

Model:

Ein Bild, das Screenshot enthält.

Automatisch generierte Beschreibung

Text to Analyze: Part in Chap1 of 20000-Leagues-Under-the-Sea.txt

Ein Bild, das Screenshot enthält.

Automatisch generierte Beschreibung