

Energy Informatics

System Design — Data Analysis

Albert-Ludwigs-Universität Freiburg

Peter Thiemann

29 Jan 2018



**UNI
FREIBURG**

The crew

- Part 1: Prof. Dr. Peter Thiemann
- Part 2: Prof. Dr. Georg Lausen
- Tutoring: Gabriel Radanne

Who are you?

What will
YOU
use programming for?

Data Analysis

- Scrutinizing large data sets
meter readings, usage statistics, connection data
- Coming up with hypotheses
- Verifying the hypotheses

Data Analysis

- Scrutinizing large data sets
meter readings, usage statistics, connection data
- Coming up with hypotheses
- Verifying the hypotheses

Information Sources and Data Conditioning

- Reading data from files, CSV, XML, spreadsheet
- Cleaning up: detecting formatting errors, removing implausible data, outliers, etc

Our goal



- Simple tools for simple data analysis
- Rehearse with small examples

First application

Analysis of voting patterns



Voting data

The German Bundestag runs two kinds of voting procedures

- anonymous vote
- vote by role call

First application

Voting data

The German Bundestag runs two kinds of voting procedures

- anonymous vote
- vote by role call

Vote by role call (namentliche Abstimmung)

- for controversial topics
- protocol registers each vote along with the name of the voter
- procedure
 - there are voting cards in three colors, blue, red, and white
 - voting cards are imprinted with name and fraction
 - cards are dropped in an urn
 - outcome is published in a table

Procedure

- For the vote by role call procedure, we can download the outcome of the votes for last few years
- Use this data to demonstrate the steps
 - forming a hypothesis
 - obtaining the data
 - cleaning up the data
 - verifying the hypothesis

Voting by role call

- <https://www.bundestag.de/abstimmung>
- only accessible inside Germany
- only available in German
- Voting data from 2009-2018 in pdf and xls
- <http://www.bundestag.de/parlament/plenum/abstimmung/liste>
- etc

Two formats

- pdf: unstructured — useless
https://www.bundestag.de/blob/404086/029a3812d1a1a63979de77b48fbbabc2/20160128_2-data.pdf
- xls: structured — can load into spreadsheet program or read with Python
https://www.bundestag.de/blob/404080/8d985dd7bac5ecff733d4b98d40a2c07/20160128_2_xls-data.xls

A look through the spreadsheet



- Is this data valid?
- What can go wrong?
- Consistency?

- Consider the columns
(ja, nein, Enthaltung, ungültig, nicht abgegeben)
that is: yes, no, abstain, invalid, not voted
- Each of them contains 0/1
- Internal consistency of one voting protocol
 - Each row should contain exactly one “1” entry in these columns
 - The columns (ja, nein, Enthaltung, ungültig, nicht abgegeben) should contain as many “1” as the summary count announces

- Load into spreadsheet program
- Make a copy (save with new name)
- Create a new work sheet
- On the new sheet create a formula that checks
 - whether each of the yes/no columns contains 0 or 1
 - that the yes/no columns contain exactly one 1
 - the above two points for each row



- Mostly the fractions enforce **the whip**, i.e., they demand that their members vote according to the party line
 - some fractions do not enforce the whip
 - for some ballots, fractions do not impose restrictions
- As the voting is public, it is much tougher to “defect” or “betray” the party by voting differently

- Mostly the fractions enforce **the whip**, i.e., they demand that their members vote according to the party line
 - some fractions do not enforce the whip
 - for some ballots, fractions do not impose restrictions
- As the voting is public, it is much tougher to “defect” or “betray” the party by voting differently

Task

- Write formulas to check that parliamentarians mostly adhere to the whip
- Which parties do / do not impose restrictions?

End Part I