

# KodeKlubben 2.0

## Øvelsesgang 1

---

Kristian Urup Olesen Larsen, Jakob Jul Elben

October 30, 2018

Økonomisk Institut, KU

# Velkommen!

Hvem er vi?

- Økonomistuderende
- RA's på Økonomisk Institut
- Arbejder typisk i Python, R, STATA eller SAS

Hvem er i?

- Det er sådan set ligemeget, vi skal bare have det sjovt og lære noget *data science*.

Alt materiale ligger på GitHub

- Åben <https://github.com/Kristianuruplarsen/kodeklubben>
- Klik på knappen *Clone or download* og på *Download ZIP*

Kristianuruplarsen / kodeklubben Private

Watch 0 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

Exercises for sociology students in kodeklubben (In danish) Edit

Manage topics

1 commit 1 branch 0 releases 1 contributor

Branch: master New pull request

Create new file Upload files Find file Clone or download

Kristianuruplarsen first

exercises/session2

slides

Help people interested in this repository understand your project by adding a README.

Clone with HTTPS Use SSH

Use Git or checkout with SVN using the web URL.

<https://github.com/Kristianuruplarsen/>

Open in Desktop Download ZIP

# Hvad skal vi lave?

- Hvad mener (folketings)partierne egentligt? Gid vi kunne spørge dem...

# Hvad skal vi lave?

- Hvad mener (folketings)partierne egentligt? Gid vi kunne spørge dem...
- DR's kandidattest fra KV17 er stadig online!

# Hvad skal vi lave?

- Hvad mener (folketings)partierne egentligt? Gid vi kunne spørge dem...
- DR's kandidattest fra KV17 er stadig online!
- Alle kandidaternes svar er "frit" tilgængelige.

# Hvad skal vi lave?

- Hvad mener (folketings)partierne egentligt? Gid vi kunne spørge dem...
- DR's kandidattest fra KV17 er stadig online!
- Alle kandidaternes svar er "frit" tilgængelige.

Kort note om formattet:

- Vi har slides der viser hvordan i kan løse alle spørgsmålene,

# Hvad skal vi lave?

- Hvad mener (folketings)partierne egentligt? Gid vi kunne spørge dem...
- DR's kandidattest fra KV17 er stadig online!
- Alle kandidaternes svar er "frit" tilgængelige.

Kort note om formattet:

- Vi har slides der viser hvordan i kan løse alle spørgsmålene, men den bedste vej frem er *learning by doing*



# Hvad skal vi lave?

- Hvad mener (folketings)partierne egentligt? Gid vi kunne spørge dem...
- DR's kandidattest fra KV17 er stadig online!
- Alle kandidaternes svar er "frit" tilgængelige.

Kort note om formattet:

- Vi har slides der viser hvordan i kan løse alle spørgsmålene, men den bedste vej frem er *learning by doing*
- 1) Læs dokumentation 2) spørg en sidemakker 3) google det (stackoverflow har som regel løsningen) 4) spørg os.

Op til KV17 lavede DR en kandidattest. Alle kandidaterne har selvfølgelig svaret på spørgsmålene. De svar vil vi gerne have.

Denne gang:

- Download data fra DR
  - internet-hacks
  - Interagere med nettet gennem python
- Rens datasættet og gør klar til alt det sjove
  - jonglere med data og formatter

Næste gang:

- Alt det sjove.
  - *Dimensionality reduction*
  - Interaktive plot

## Øvelser

---

## Opgave 1.1

Det link i leder efter er:

*<https://www.dr.dk/tjenester/kv17-candidateapi/api/constituency/all>*

Prøv at åbne linket i en browser!

## Opgave 1.2

- Hvis ikke i har installeret requests er det en god ide at gøre det nu
- Google:
  - “python requests get link”
  - “python requests json”

```
1 import requests
2
3 url = #[FILL IN]
4
5 response = requests.get(url)
6 kommuner = response.json()
7
```

## Opgave 1.3

- Kig i den json fil i har downloadet - hvordan er den struktureret?
- I kan hente alle slugs med et loop:

```
1 slugs = list()
2 for x in kommuner['constituencies']:
3     if not x['slug'] == 'ikke-oplyst':
4         slugs.append(x['slug'])
5
```

- Eller med en *list comprehension*:

```
1 slugs = [x['slug'] for x in kommuner['constituencies']
2           if not x['slug'] == 'ikke-oplyst']
3
```

# Opgave 1.4

- I skal ind i jeres developer console igen - denne gang leder i en json fil med samme navn som den kommune i har søgt på

The screenshot shows the Chrome DevTools Network tab. The list of requests includes:

Sts...	Meth...	File	Domain	Cause	Type	Transferr...	Size	...
200	GET	all	www.dr.dk	fetch	json	2.06 KB	11.77 KB	15 ms
200	GET	all	www.dr.dk	fetch	json	2.06 KB	11.77 KB	20 ms
200	GET	expand-data.json	www.dr.dk	xhr	json	1.03 KB	2.14 KB	7 ms
200	GET	me	www.dr.dk	xhr	json	642 B	114 B	70 ms
200	GET	all	www.dr.dk	fetch	json	2.06 KB	11.77 KB	53 ms
200	GET	kalundborg	www.dr.dk	fetch	json	4.53 KB	15.13 KB	43 ms
200	GET	campaigns.json?edps...	www.dr.dk	xhr	json	1.01 KB	261 B	79 ms
200	GET	is7c584a89be4e5a1df...	pp.lp4.io	img	plain	134 B	0 B	36 ms
200	GET	p7li584a89be4e5a1df...	pp.lp4.io	img	plain	134 B	0 B	38 ms
200	GET	p7li584a89be4e5a1df...	pp.lp4.io	img	plain	134 B	0 B	37 ms
200	GET	u7li584a89be4e5a1df...	pp.lp4.io	img	plain	115 B	0 B	185 ms
200	GET	12087-Hans-Munk-a.pn...	asset.dr.dk	img	png	98.49 KB	98.20 KB	17 ms
200	GET	anonymous.svg	www.dr.dk	img	svg	1.37 KB	1.47 KB	10 ms
304	GET	favicon.ico	www.dr.dk	img	x-icon	cached	1.12 KB	8 ms
200	GET	/DR.Common.Banner.A...	www.dr.dk	xhr	xml	600 B	100 B	57 ms
200	GET	package	www.dr.dk	xhr	xml	473 B	0 B	8 ms

The selected request 'kalundborg' has the following response properties:

```
{
  slug: kalundborg,
  greaterConstituencyId: 1,
  ballotParties: [...],
  partyId: 1,
  partyName: Socialdemokratiet,
  partyLetter: A,
  candidates: [...],
  id: 11473,
  firstName: Gunver,
  lastName: Jensen,
  photoUrl: http://www.altinget.dk/images/kv17/kvadrot-11473-gunverjensen-a-721.jpg,
  partyLetter: A,
  partyName: Socialdemokratiet,
  isPartyLeader: false,
  answerString: 11125531144343
}
```

## Opgave 1.4

```
1 def hent_kommune_data(kommuneslug):
2     base = 'https://www.dr.dk/tjenester/kv17-candidateapi/'
3     url = base + f'api/ballot/{kommuneslug}'
4
5     response = requests.get(url)
6
7     if response.ok:
8         return response.json()
9     else:
10        return None
11
```



# Opgave 1.5

```
1  import pandas as pd
2
3  def rens_kommune_data(kommunejson):
4      ballot = kommunejson['ballot']['constituency']
5      parties = kommunejson['ballot']['ballotParties']
6
7      _name = ballot['_slug']
8      _id = ballot['_id']
9
10     for party in parties:
11         party_candidates = party['_candidates']
12
13         for candidate in party_candidates:
14             _fname = candidate['_firstName']
15             _lname = candidate['_lastName']
16
17             data = {'id': [_id],
18                    'kommune': [_name],
19                    'letter': [candidate['_partyLetter']],
20                    'party': [candidate['_partyName']],
21                    'name': ['{} {}'.format(_fname, _lname)],
22                    'answers': [candidate['_answerString']]
23                }
24
25         try:
26             df = pd.concat([df, pd.DataFrame(data)])
27         except NameError:
28             df = pd.DataFrame(data)
29
30     return df.reset_index(drop = True)
```

## Opgave 1.6

```
1 from time import sleep
2
3 for kom in slugs:
4     sleep(2)
5     print(kom)
6     data = hent_kommune_data(kom)
7     clean = rens_kommune_data(data)
8
9     try:
10         df = pd.concat([df, pd.DataFrame(clean)])
11     except NameError:
12         df = pd.DataFrame(clean)
```

## Opgave 1.7

```
1 df.reset_index(drop = True)\  
2   .to_csv('../data/candidates.csv',  
3           index = False)  
4
```

**Næste gang/tak for i dag**

---

## Næste gang/tak for i dag

- Vi lægger en rettevejledning ud til dagens øvelser på github
- Vi lægger datasættet op før næste øvelsesgang - *ikke* nødvendigt med hjemmearbejde!

## Næste gang/tak for i dag

- Vi lægger en rettevejledning ud til dagens øvelser på github
- Vi lægger datasættet op før næste øvelsesgang - *ikke* nødvendigt med hjemmearbejde!

Næste gang:

- Samme tid, samme sted, næste fredag.
- Vi skal lave noget analyse på de data i har scrapet i dag.