



Data science and public administration

Konference om digitale modeller i samfundsvidenskabelig forskning –
IT Vest

26. august 2021

Matt W. Loftis



Who am I?

Matt W. Loftis

Associate professor

AU Political science

Public administration section

- ▶ Political accountability
- ▶ Control of bureaucracy
- ▶ Political agenda
- ▶ Applied digital & computational methods



Digital methods in my work

“big” data on the public sector

1. *Build* data sets from:
 - ▶ many local gov web sites
 - ▶ national gov data bases
 - ▶ media/IO/regulators, etc.
2. *Applied ML* to measure concepts
 - ▶ label text
 - ▶ find patterns
3. Statistical models to test hypotheses

Digital methods in my teaching



Teaching & speaking:

- ▶ “Political Data Science”
 - ▶ Automated data collection
 - ▶ Data visualization
 - ▶ Data management
 - ▶ Applied ML
 - ▶ Applied observational statistics
- ▶ Summer schools & workshop



Data *collection* = distinct skill to learn

The public sector internet is huge

- ▶ In 2011, the Obama admin launched a campaign to cut waste by reducing the number of US federal government websites
- ▶ At the time, that was $\approx 2,000$ full websites
- ▶ Those included $\approx 24,000$ smaller sub-sites
- ▶ <https://obamawhitehouse.archives.gov/blog/2011/06/13/toomanywebsitesgov>



Upshot of digitalization

Why invest in digital/computational skills?

- ▶ Fast (eventually)
- ▶ Thorough
- ▶ Integrates
- ▶ Low budget (possibly)

Why academic researchers in particular?

- ▶ Shrinking budgets
- ▶ Information growth
- ▶ Information decay
- ▶ Expectation *growth*

How I ended up here



- ▶ Background as anti-corruption activist
- ▶ Came to academia to understand why we fail!



How I ended up here

- ▶ Background as anti-corruption activist
- ▶ Came to academia to understand why we fail!

What I found

- ▶ Started in poli sci \approx 2009
- ▶ Comparative politics was light on data
- ▶ *Laws* got all the attention
- ▶ I needed to get more data. . .



Example applications

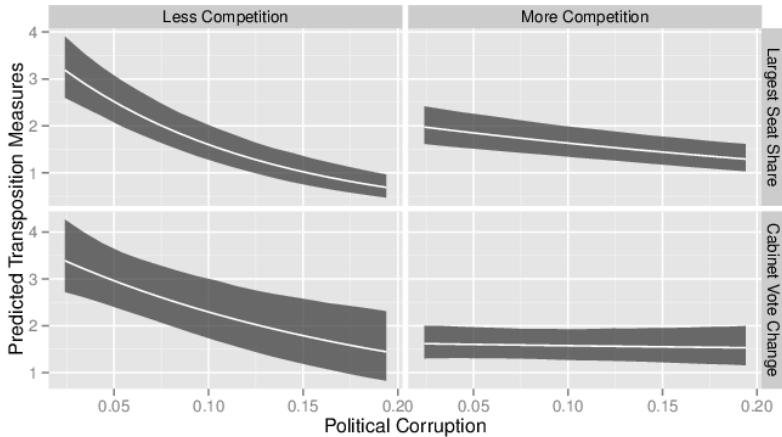
- ▶ Politicians delegate to bureaucracy to avoid blame
- ▶ Corruption + informal influence over bureaucracy \Rightarrow more delegation

Data

- ▶ 1000s of transposition reports from EU's Eur-Lex
- ▶ 100s of Commission directives
- ▶ Downloaded using bots
- ▶ Processed automatically



Example applications





Example applications

- ▶ Local government agendas
- ▶ Which political issues do Danish local councils talk about?
And why?

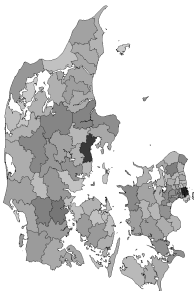
Data

- ▶ Byråd dagsorden
- ▶ 98 kommuner, 2007-2016
- ▶ +200,000 agenda items
- ▶ ML classifier to label policy topics

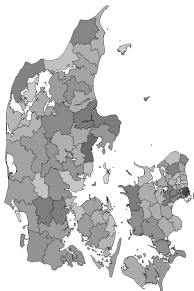


Example applications

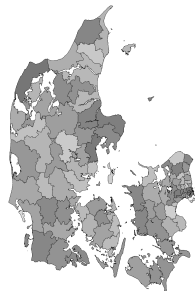
Agenda points



Subcategories

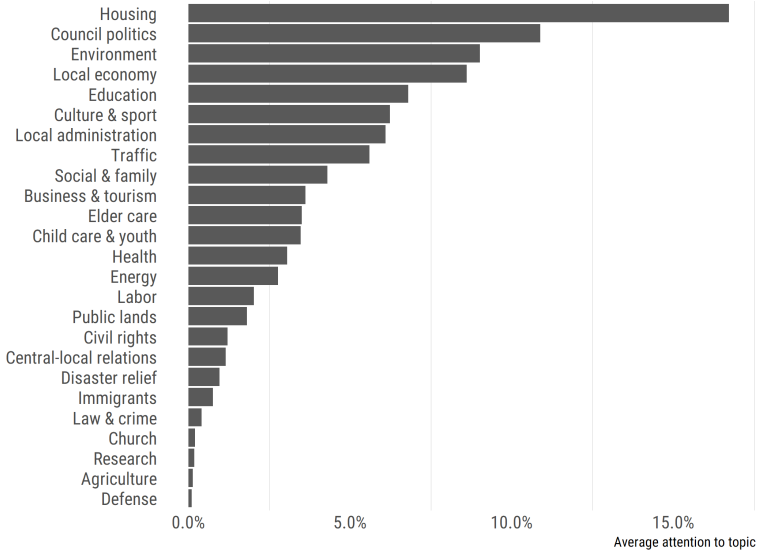


Agenda complexity





Example applications



My path developing skills



- ▶ Entered grad school at 0 (zero) tech skills
- ▶ Began with collecting data at scale
- ▶ Lots of data \Rightarrow measurement difficulty
- ▶ Applied ML let me label data and more



Lessons were learned(?)

Subjective perspective on doing **applied** work:

- ▶ Be ruthlessly *problem-driven*
(Learn new skills to do something)



Lessons were learned(?)

Subjective perspective on doing **applied** work:

- ▶ Be ruthlessly *problem-driven*
(Learn new skills to do something)
- ▶ Coding languages are '*languages*'
(Languages aren't a subject in themselves)



Lessons were learned(?)

Subjective perspective on doing **applied** work:

- ▶ Be ruthlessly *problem-driven*
(Learn new skills to do something)
- ▶ Coding languages are '*languages*'
(Languages aren't a subject in themselves)
- ▶ Learn basic principles for coding/debugging
(Commenting, version numbering, etc. also git!)

What would I do differently?



- ▶ Follow my own advice :)
- ▶ Take smaller bites

Thank you!



Feel free to get in touch!
mattwloftis@ps.au.dk