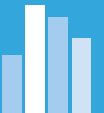


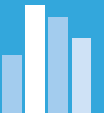
Session 3.1

- Danish experiences with Web-scraping



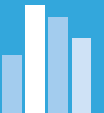
Other data integration projects

- Applied for grant EU grant "04151.2014.002-2014.420" considering web scraping
- Ended August 2016



Procedure

1. How is web scraping carried out in other countries
2. Which IT-software can be used
3. Are there any legal obstacles?
4. Tests
5. Does it save any costs?
6. Possible implementation



Scraping period 1

- We opted to use the free software of Import IO. This was chosen to avoid putting many IT resources into building our own web scrapers, and it was free and didn't create any legal obstacles
- Self-scraping:
- Period from mid 2015-mid 2016
- We scraped 5-10 pages, once per week
- We used the program Import IO (free then, not now!)

Conclusion on tests

- With the current number of observations collected from each internet vendor, we do not at the moment see a potential in switching to web scraping.
- **Why?**
- There is a lot of info missing on several variables. This makes quality corrections difficult
- We could create a bias in only collecting from a few vendors.
- The layout changes of web-pages makes it only efficient to scrape from big vendors



- Outsourcing:
- Period of first half of 2018
- Carried out by Hitachi Denmark for a fee of a few thousand DKK per month per group
- Done on 2 homepages that have national pricing policy in general including on TVs and cell phones
- Intended use: hedonic models!

- Results: **Mixed!**
- They were not able to do the scraping in a fixed format

