1: Introduktion

Videregående kvantitative metoder i studiet af politisk adfærd

Frederik Hjorth
fh@ifs.ku.dk
fghjorth.github.io
@fghjorth

Institut for Statskundskab
Københavns Universitet

6. september 2017

Mig:

- cand.scient.pol., ph.d. i statskundskab
- post.doc. ved IfS, ansat på projektet *Digital Disinformation*
- interesser: politisk psykologi, holdningsdannelse, indvandring, anvendt metode

Jer:

- navn
- ønsker til faget
- god tekst jeg har læst under studiet

- "Videregående Kvantitative Metoder I Studiet Af Politisk Adfærd" (VKM)
- Seminar m. 28 timers holdundervisning
- Hjemmeside: `github.com/fghjorth/vkme17`

Fagets opbygning

Blok 1

Gang	Tema	Litteratur	Case
1	Introduktion til R	Leeper (2016)	
2	R workshop + tidy data	Wickham (2014), Zhang (2017)	
3	Regression I: OLS brush-up	AP kap 3	Newman et al. (2015), Solt et al. (2017)
4	Regression II: Paneldata	AGS kap 4	Larsen et al. (2016)

Fagets opbygning

Blok 2

5	Introduktion til kausal inferens	Hariri (2012), Samii (2016)	
6	Matching	Justesen & Klemmensen (2014)	Nall (2015)
Efterårsferie			
7	Eksperimenter I	AP kap 1, GG kap 1+2	Gerber, Green & Larimer (2008)
8	Eksperimenter II	GG kap 3+4+5	Gerber & Green (2000)
9	Instrumentvariable	AP kap 4	Lundborg et al. (2017)
10	Difference-in-differences	AP kap 5	Enos (2016)
11	Regressionsdiskontinuitetsdesigns	AP kap 6	Eggers & Hainmueller (2009)

Fagets opbygning

Blok 3

12	Tekst som data	Grimmer & Stewart (2013), Benoit & Nulty (2016)	Baturo & Mikhaylov (2013)
13	Scraping af data fra online-kilder	MRMN kap 9+14	Hjorth (2016)
14	'Big data' og maskinlæring	Varian (2014), Montgomery & Olivella (2017)	Theocharis et al. (2016)

Pensum

- AGS: Andreß, H. J., Golsch, K., & Schmidt, A. W. (2013). Applied panel data analysis for economic and social surveys. Springer Science & Business Media.
- GG: Gerber, A. S., & Green, D. P. (2012). Field experiments: Design, analysis, and interpretation. WW Norton.
- AP: Angrist, J. D., & Pischke, J. S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton university press.
- MRMN: Munzert, S., Rubba, C., Meißner, P., & Nyhuis, D. (2014). Automated data collection with R: A practical guide to web scraping and text mining. John Wiley & Sons.

+ hertil artikler og case-artikler

Målbeskrivelse Seminarets målsætning er at sætte den studerende i stand til efter endt undervisning at kunne (ILO's):

- Identificere relevante designs og teknikker for at løse politologiske problemstillinger.
- Bearbejde data i strukturerede og ustrukturerede formater mhp. senere analyse.
- Analysere empiriske politologiske problemstillinger med udgangspunkt i kvantitative data.
- Reflektere over fordele og ulemper ved forskellige designs og teknikker fra kursets pensum og samt i andres og eget arbejde.

Eksamen

Formelle krav:

- seminaropgave på 10-20 ns.
- skal skrives individuelt
- eksamen forudsætter mindst 75 pct. tilstedeværelse
- aflevering af oplæg til seminaropgave fredag d. 27/10 kl. 16 (tbc)

Koncepter for opgaven:

1. Fri opgave med anvendelse af fagets metoder
2. Replikationsstudie
3. Specialeforstudie

3 'ben' i faget:

1. Logik: styrker og svagheder ved forskellige undersøgelsesdesigns
2. Teknik: den underliggende økonometri/statistik
3. Implementering: hvordan man faktisk gennemfører analysen i R

Fokus i VKM på 1+3

Typisk struktur for holdtime: iht. "Particular General Particular" princippet

- Motiverende eksempel på metode
- Præsentation af principper
- Gennemgang af implementering i R

Bærende motivation for faget: 2 revolutioner har drevet kolossal vækst i kvantitativ samfundsvidenskab

1. 'data revolution'
2. 'computational revolution'

(OBS: i uge 5 møder vi en tredje, den såkaldte 'credibility revolution')

Konsekvens: fortrolighed med velstrukturerede data er utilstrækkeligt

»It is simply not sufficient to achieve 'statistical literacy' by learning about common statistical concepts and methods. Instead, all students in the social sciences should acquire basic data analysis skills so that they can exploit ample opportunities to learn from data (...)« (Kosuke Imai: *Quantitative Social Science: An Introduction*, p. 3)

Hvad er R?

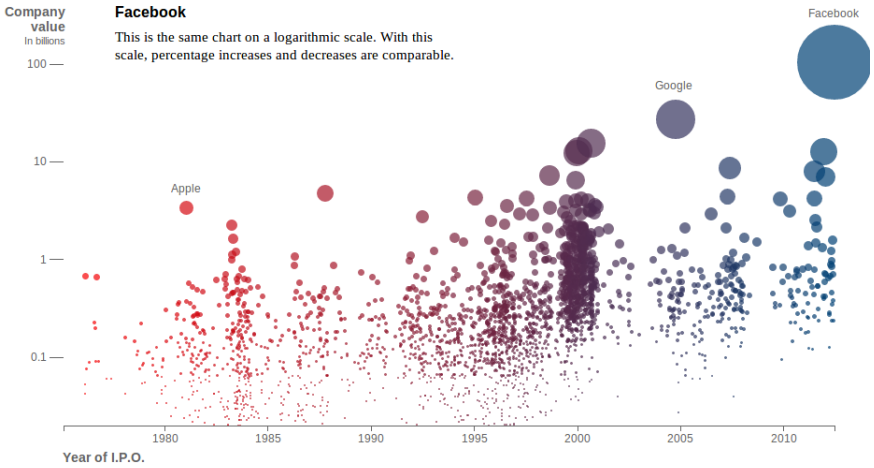
- et program til statistisk programmering
- et programmeringssprog (som C++, Python, Perl, etc.)
- fungerer generelt *objekt-orienteret* (ctr. fx. Stata)
- open source
- opfindere: Ross Ihaka & Ross Gentleman
- videreudvikling af *S*

Første spadestik til R for ca. 20 år siden:

→ første stabile beta lanceres i 2000

Hvorfor R?

- næsten uendelige anvendelser
- reproducerbart workflow
- den nye analytiske standard
- free as in free speech and free beer
- absolut bedst til datavisualisering

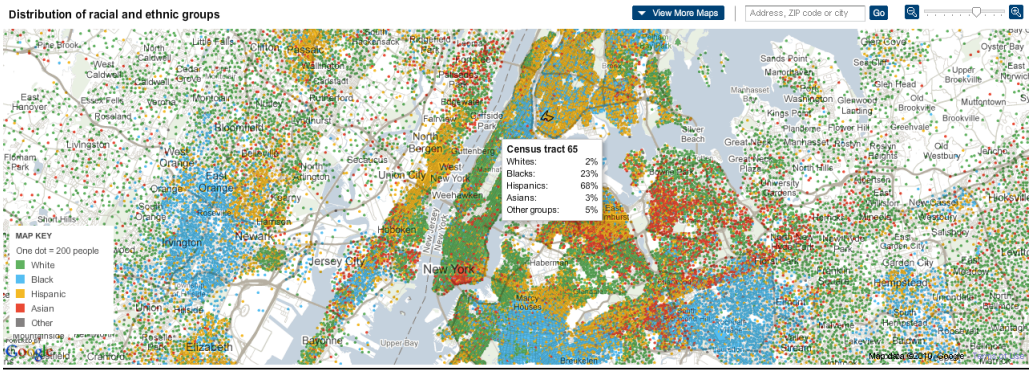

Company value
In billions
Facebook
This is the same chart on a logarithmic scale. With this scale, percentage increases and decreases are comparable.
Facebook
Google
Apple
100
10
1
0.1
1980
1985
1990
1995
2000
2005
2010
Year of I.P.O.

The New York Times

Mapping America: Every City, Every Block

Browse local data from the Census Bureau's American Community Survey, based on samples from 2005 to 2009.

Find something interesting? Share this view on Twitter or Facebook



By MATTHEW BLOCH and SHAN CARTER | Source: 2005-9 American Community Survey, Census Bureau, socialexplorer.com

Hvorfor *ikke* R?

- ingen grafisk brugerflade (GUI)
- konstant import af ekstrapakker
- meget følsom over for fejl
- kryptiske fejlmeddelelser

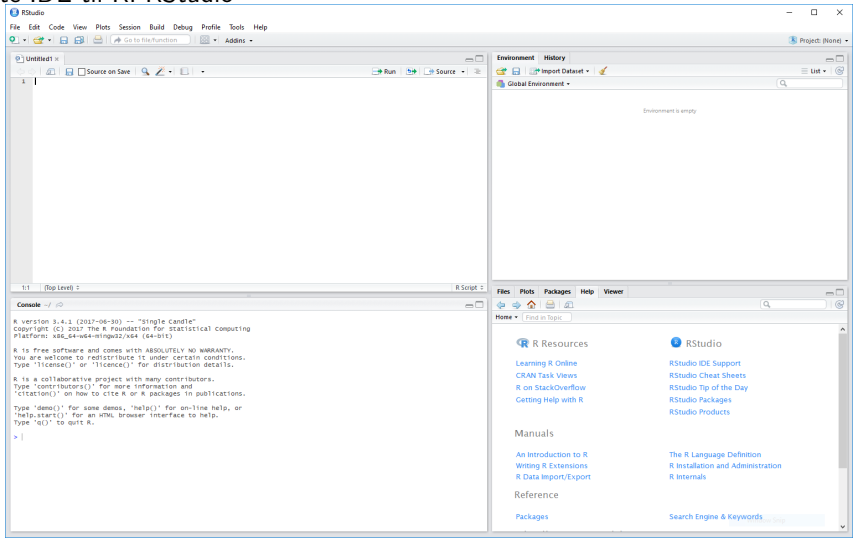
Men:

I Was Intimidated by Coding Until I Learned This Secret Strategy: Googling

"You don't need to go to grad school. Save your money. I'll teach you how to code." Seven years ago, in a bar near downtown Los Angeles, I was sharing ...

SLATE.COM

Det bedste IDE til R: RStudio



Leeper, *Really Introductory Introduction*:

- Getting started
  - brug af R som regnemaskine: fx. `(2+4)/7`
  - parsing errors ctr. syntax errors
  - nye vektorer: fx. `dice <- c(2,2,3,4)`
  - ekstrahering fra vektorer: fx. `dice[1:3]`
  - ny data frame: fx. `df <- data.frame(dice,number=1:4)`
  - data framens struktur: `str(df)`
  - centrale tendenser: `summary(df)`

- Real data
  - installer pakker: `install.packages()`
  - indlæs pakker: `library()`
  - importér data: `import()` fra `rio`-pakken

- Randomness
  - sample fra en vektor: `sample()`
- Plots
  - pakke: `ggplot2`
  - fx. `ggplot(iris,aes(x=Sepal.Length)) + geom_histogram`

- Basic programming tools
  - funktioner: fx. `ftoc <- function(f){ c<-((f-35)*5)/9 ; print(c) }`
  - for loops: fx. `for (i in 1:10) print(i*i)`

Næste gang:

- intro til 'tidy data': Wickham
- databehandling med tidyverse: Zhang
- lektie:
  - `install.packages("swirl")`
  - `library("swirl")`
  - download kursus nr. 1
  - gennemgå modul 1-4 (eller mere)

Tak for i dag!