

Opsamling
○

Tidying
○○○

Visualisering
○○○○○○○○○○○○

Tips & Tricks
○

Øvelse
○○

Kig fremad
○○

7: Workshop II

Videregående kvantitative metoder i studiet af politisk adfærd

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

Institut for Statskundskab
Københavns Universitet

25. oktober 2018

1 Opsamling

2 Tidying

3 Visualisering

4 Tips & Tricks

5 Øvelse

6 Kig fremad

Sidste gang:

- konceptuelt om paneldata
- lang vs. bred form
- omitted variable bias i paneldata
- håndtering af autokorelation
- case: Mutz om status threat (forts.)

dirty_data.xlsx:

	A	B	C	D	E	F	G	H	I	J	K	L
1	First Name	Last Name	Employee Status	Subject	Hire Date	% Allocated	Full time?	do not edit! ---->	Certification	Certification	Certification	
2	Jason	Bourne	Teacher	PE	39690	75%	Yes		Physical ed	Theatre		
3	Jason	Bourne	Teacher	Drafting	39690	25%	Yes		Physical ed	Theatre		
4	Alicia	Keys	Teacher	Music	37118	100%	Yes		Instr. music	Vocal		
5	Ada	Lovelace	Teacher	#REF!	27515	100%	Yes		PENDING	Com		
6	Desus	Nice	Administration	Dean	41431	100%	Yes		PENDING			
7	Chien-Shiung	Wu	Teacher	Physics	11037	50%	Yes		Science 6-12	Physics		
8	Chien-Shiung	Wu	Teacher	Chemistry	11037	50%	Yes		Science 6-12	Physics		
9												
10	James	Joyce	Teacher	English	32994	50%	No			English 6-12		
11	Hedy	Lamarr	Teacher	Science	27919	50%	No		PENDING			
12	Carlos	Boozer	Coach	Basketball	42221	#N/A	No		Physical ed			
13	Young	Boozer	Coach	TBD	34700	#N/A	No			Political sci.		
14	Micheal	Larsen	Teacher	English	40071	80%	No		Vocal music	English		
15												
16												

Sam:
Put alternate
certifications in
columns to right

Uundværlig pakke: janitor

- rens navne m. `clean_names()`
- fjern tomme rækker og kolonner m. `remove_empty()`
- lav tal om til datoer m. `excel_numeric_to_date()`
- → naturligvis helt igennem tidyverse-kompatibel



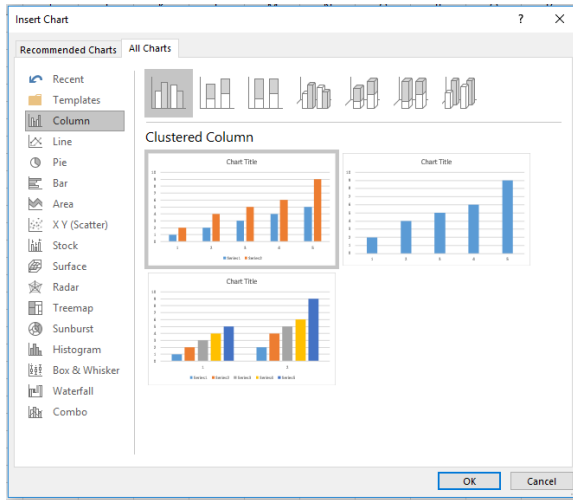
To overordnede metoder:

- enkel omkodning m. `ifelse()`
- til mere komplicerede omkodninger: `case_when()`

Visualisering af data:

- rigtig god konceptuel gennemgang: <http://socviz.co>
- her fokus på teknisk implementering m. ggplot2
- kilde: http://jcyhong.github.io/ggplot_demo.html

Datavisualisering i Excel:



Opsamling

○

ggplot2

Tidying

○○○

Visualisering

○○●○○○○○○○○

Tips & Tricks

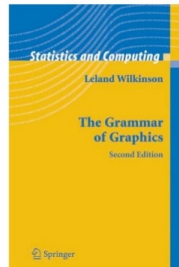
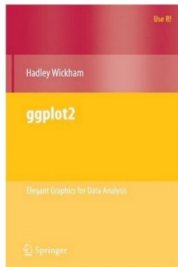
○

Øvelse

○○

Kig fremad

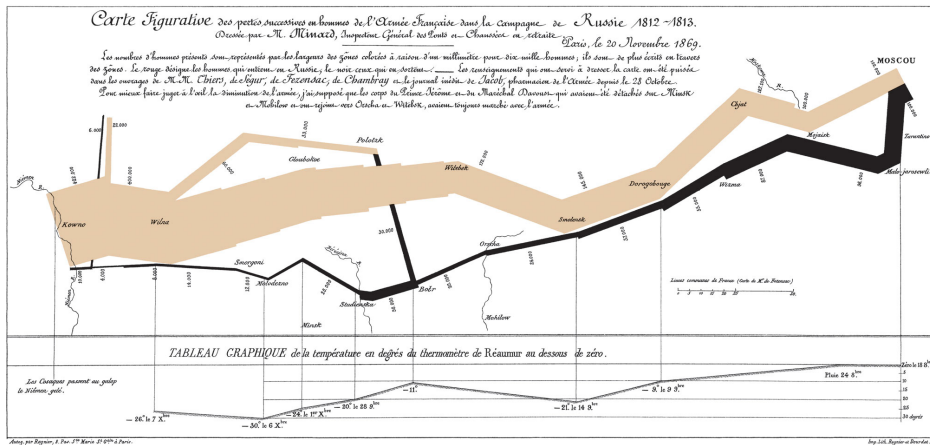
○○



'grammar of graphics' → tre centrale dele i et plot:

- ① *data*
- ② *aesthetic mappings* af variable til visuelle objekter
- ③ *geoms*, dvs. forskellige typer repræsentationer af objekterne

→ 'grammatikken' muliggør fleksibel kombination af *geoms*, opdeling i miniplots (*facets*), m.m.



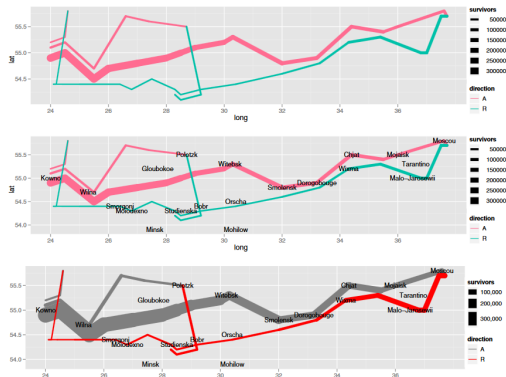
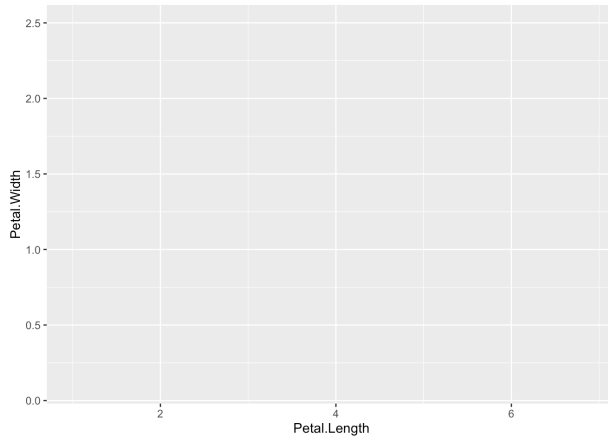


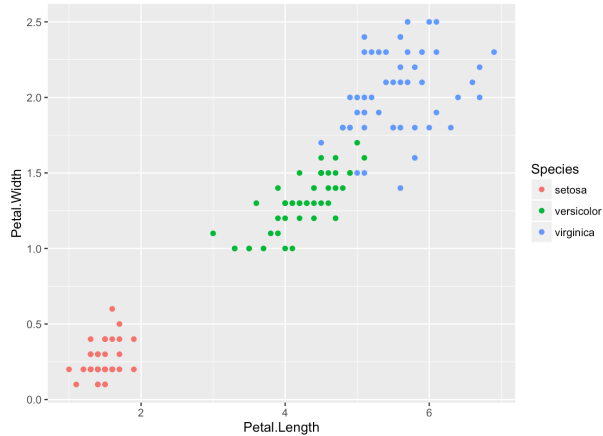
Figure 12. Iteratively reproducing the depiction of Napoleon's March by Minard. (Top) Displaying the key troop movement data. (Center) Adding town locations as reference points. (Bottom) Tweaking scales to produce polished plot.

Wickham, H. (2010). A layered grammar of graphics. *Journal of Computational and Graphical Statistics*, 19(1), 3-28.

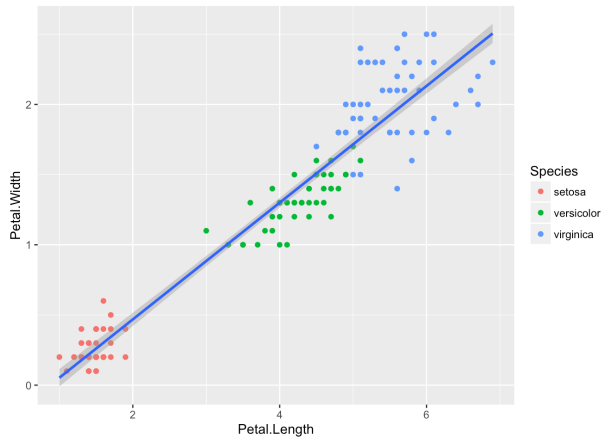
```
p1 <- ggplot(data = iris, aes(x = Petal.Length, y = Petal.Width))  
p1
```



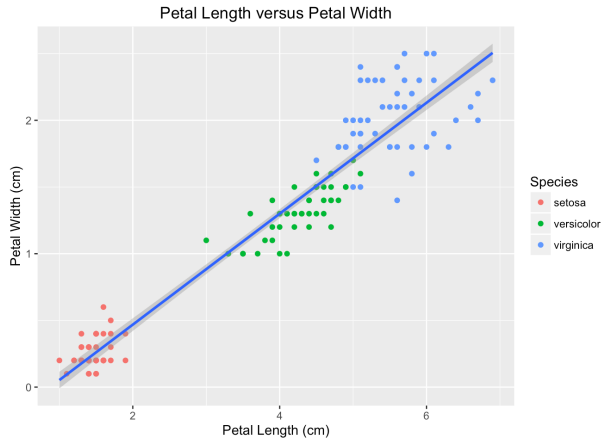
```
p2 <- p1 + geom_point(aes(color = Species)) p2
```



```
p3 <- p2 + geom_smooth(method='lm') p3
```




```
p4 <- p3 + xlab("Petal Length (cm)") + ylab("Petal Width (cm)") + ggtitle("Petal Length versus Petal Width") p4
```



ggplot2 workshop: <https://tutorials.iq.harvard.edu/R/Rgraphics/Rgraphics.html>

- kød kode frem til aktuel linje: **Ctrl** + **B**
- flere markører på samme tid: **Ctrl** + **Alt** + **klik**
- flyt linje op og ned: **Alt** + **↕**
- kopiér linje op og ned: **⇧** + **Alt** + **↕**

Tema: hvorfor vandt Leave?



Opgaver:

- 1 indlæs den britiske udgave af ESS bølge 8 (07_ESS8GB.dta) og det landespecifikke survey (07_ESS8GB_cs.dta)
- 2 merge de to datasæt med variablen idno
- 3 estimér en model der forklarer variablen leavevote vha. div. uafhængige variable
- 4 beregn den forudsagte sammenhæng mellem udvalgte variable og leavevote vha. `ggeffects::ggpredict()`
- 5 visualiser sammenhængen m. `ggplot()`

Hints:

- pdf med oversigt over variable i data-mappen
- bud på interessante prædiktorer: `ppltrst`, `trstep`, `stfec`, `stfhlth`, `imueclt`, `atcherp`, `clmchn`, `eduagb2`, `uemp3m`, `imptrad`

Næste gang:

- konceptuel introduktion til kausal inferens og 'credibility-revolutionen'
- Angrist & Pischke (2010): god kontekst om credibility-revolutionen, læs evt. kursorisk, spring s. 18-22 over
- Samii (2016): god indføring i revolutionen med politologi-briller – læs dog afsnittet om 'pseudo-generality' kursorisk
- case-tekst: Carroll (2018) → NYTimes-artikel, sjov illustration af temaet

Opsamling
○

Tidying
○○○

Visualisering
○○○○○○○○○○○○

Tips & Tricks
○

Øvelse
○○

Kig fremad
○●

Tak for i dag!