

SOK 1004

Økonomiske emner og programmering

Forelesning 3: tidyverse og ggplot2

Even Soltvedt Hvinden

UiT Norges arktiske universitet

`even.c.hvinden@uit.no`

Introduksjon

- **Mål:** Bruke grunnleggende kommandoer fra dplyr og ggplot2 til å lage figurer
- **Transformere data:** Fem kommandoer i dplyr:
`select()`, `filter()`, `arrange()`, `mutate()`,
`summarize()`
- **Visualisere data:** Vi bruker ggplot2: i kombinasjon med dplyr
- **Gjennomføring:** Jeg introduserer kommandoene, dere prøver selv ved hjelp av vedlagt kode

koden vi skal bruke i forelesningen

Lagre data i en tibble

- En tibble er det anbefalte lagringsformatet i tidyverse
- **Eksempel:** se co2data i koden til forelesningen

	iso_code	year	...	co2
	<chr>	<dbl>	...	<dbl>
1	AFG	1949	...	0.015
2	AFG	1950	...	0.084
⋮	⋮	⋮	⋮	⋮
25989	ZWE	2020	...	10.5

Velg variable med `select()`

- Bruk `select()` til å velge variable (kolonner) som egen tabell
- **Eksempel:** `select(co2data, iso_code)`

	iso_code
	<chr>
1	AFG
2	AFG
⋮	
25989	ZWE

Velg verdier med filter()

- Bruk `filter()` til å velge verdier (rader) som egen tabell
- **Eksempel:** `filter(co2data, iso_code == "AFG")`

	iso_code	year	...	co2
	<chr>	<dbl>	...	<dbl>
1	AFG	1949	...	0.015
2	AFG	1950	...	0.084
⋮	⋮	⋮	⋮	⋮
60	AFG	2020	...	12.2

Bruk ggplot2 til å lage figurer

- **Tre valg:** Data – aes() (koordinatsystem) – og geom_ (visualisering, farger, linjestørrelse) \implies figur
- **Råd:** Bruk cheat sheet og pipes

```
co2data %>%  
filter(country == "Norway") %>%  
ggplot(aes(x=year, y=co2)) %>%  
+ geom_line() %>%  
+ theme_minimal()
```

Nå: Bruk RStudio til å gjøre oppgave 1 - 4 i koden til forelesning 3

Endre rekkefølge på verdier med `arrange()`

- Bruk `arrange()` til å endre rekkefølgen på verdier
- **Eksempel:** `arrange(co2data, co2)`

	iso_code	year	...	co2
	<chr>	<dbl>	...	<dbl>
1	ARM	1830	...	0
2	ARM	1831	...	0
⋮	⋮	⋮	⋮	⋮
25989	na	1857	...	na

Lag nye variable med `mutate()`

- Bruk `mutate()` til å lage en ny variabel
- **Eksempel:** `mutate(co2data, x = year + 1)`

	iso_code	year	x	...	co2
	<chr>	<dbl>	<dbl>	...	<dbl>
1	AFG	1949	1950	...	0.015
2	AFG	1950	1951	...	0.084
⋮	⋮	⋮	⋮	⋮	⋮
25989	ZWE	2020	2021	...	10.5

Få deskriptiv statistikk med `summarize()`

- Bruk `summarize()` til å få en tabell med deskriptiv statistikk, eksempelvis sum, gjennomsnitt
- **Eksempel:** `summarize(group_by(co2data, year), tot_co2 = sum(co2))`

	year	tot_co2
	<dbl>	<dbl>
1	1750	56.1
2	1751	56.1
⋮	⋮	⋮
271	2020	NA

Nå: Bruk RStudio til å gjøre oppgave 5 - 8 i koden til forelesning 3