



Grunnkurs

DAG 2

SUSIE JENTOFT, ASLAUG FOSS, TATSIANA PEKARSKAYA



Statistisk sentralbyrå
Statistics Norway

Mål

- Kjennskap til programmet R og RStudio
- Åpne RStudio og kjøre enkle beregninger
- Lese inn data
- Behandler data
- Lage tabeller og oppsummere
- Lage figurer



Oppsummering fra dag 1

- Skriv kode i RStudio som source fil og kjøre med ctrl + enter
- Lage objekter ved: `<-`
- Lage vektorer ved: `c()`
- Kalle biblioteker/tilleggpakker med `library()`
- Les inn data med `read_csv2()` eller `read_sas()`



R i andre sone

- Linux:
 - Start linux fra PC med å laste ned Secure Global Desktop Client (Linux) fra Programvaresenter.
 - R og RStudio er på : sl-sas-compute-01/02/03/04 (SAS Terminal), og sl-stata-p3
 - Skriv «rstudio» i terminal
- Prod. sone windows – start RStudio fra windows-meny
- Dapla – R i jupyter
- <https://wiki.ssb.no/display/s880/For+R+brukere>



Data behandling med tidyverse

- Gjør koden ryddigere
- Pipelines %>%

Base R:

```
leave_house(get_dressed(get_out_of_bed(wake_up(me))))
```

tidyverse:

```
me %>%  
  wake_up() %>%  
  get_out_of_bed() %>%  
  get_dressed() %>%  
  leave_house()
```

Lage nye variabler: mutate()

- Kan brukes som en del av en pipeline

```
datanavn %>%  
  mutate(nyvariabel = 1000)
```

Gir variabel et navn

```
datanavn %>%  
  mutate(nyvariabel = oldvariabel * 1000)
```

Gir variabel et navn

Eksisterende variabel

+ hva skal gjøres



Statistisk sentralbyrå
Statistics Norway

Lage nye variabler: mutate()

- Kombinere med `ifelse()`
- Variablene ikke lagres hvis ikke `<-` brukes
- Kan overskrives
- Flere variablene kan lagres samtidig (, for å skille)
- Endre variabeltype (`as.character()`, `as.numeric()`)

```
datanavn %>%  
  mutate(variabelnavn = as.character(variabelnavn))
```



Velg noen rader: filter()

- For å velge ut noen rader bruker vi filter()
- Skriv logiske setning inn i parentes.
- Flere logiske setninger kan brukes sammen (skille med ,)

```
datanavn %>%  
  filter(condition)
```

- Igjen: Ingenting lagres uten <-

Velg ut noen variabler: select()

- Brukes med pipelines
- Skriv variabelnavn i parentes
- En eller flere variabler (skille med ,)
- Brukes sammen med andre funksjoner (for eks. filter())

```
datanavn %>%  
  select(variabelnavn)
```

```
datanavn %>%  
  filter(condition) %>%  
  select(variabelnavn)
```

Oppsummering/aggregering: summarise()

- Ta oppsummering (summen, gjennomsnitt, median, antall) av en variabel med summarise()

```
datanavn %>%  
  summarise(oppsummeringsnavn = mean(variabelnavn))
```

Gir oppsummering et navn

Hva skal gjøres:

- mean()
- median()
- sum()
- n()

Eksisterende variabel

Gruppering: group_by()

- Gjøre alle prosesser etterpå innen hver gruppe

```
datanavn %>%  
  group_by(grupperingsvariabel) %>%  
  summarise(oppsummeringsnavn = mean(variabelnavn))
```



Gruppering: group_by() og spread()

- Kombinere flere variabler med ,
- For en 2 x 2 frekvenstabell:

```
datanavn %>%  
  group_by(grupperingsvariabel1, grupperingsvaraibel2) %>%  
  summarise(oppsummeringsnavn = n()) %>%  
  spread(grupperingsvariabel1, oppsummeringsnavn)
```

Endre variabelnavn: rename()

```
datanavn %>%  
  rename(nyttnavn = gammeltnavn)
```

Øvelser 3

- Gå inn til samme møte/chatrommet som tidligere.
- Øvelsene til oppgavesett 1 er på fil: **øvelser_3.R**



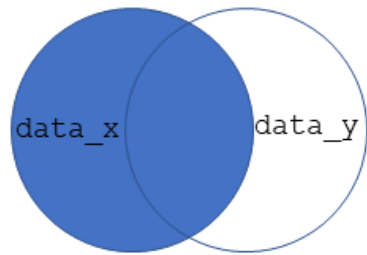
Legg til en rad: `add_row()`

- Rad må ha samme antall og type data som datasett

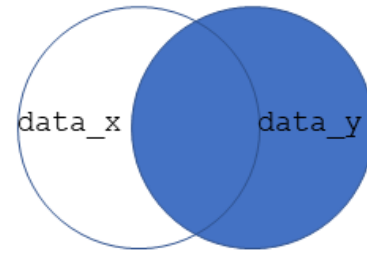
```
datanavn %>%  
  add_row(variabelnavn1 = "Oslo", variabelnavn2 = 57733)
```



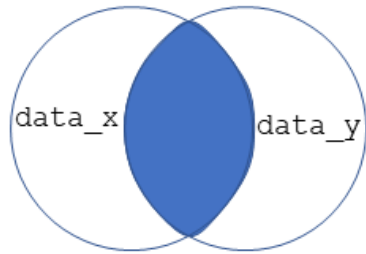
Koble to datasett



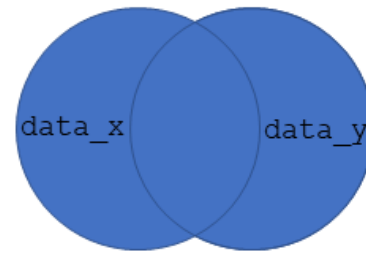
`left_join(data_x, data_y)`



`right_join(data_x, data_y)`



`inner_join(data_x, data_y)`



`full_join(data_x, data_y)`



Koble to datasett

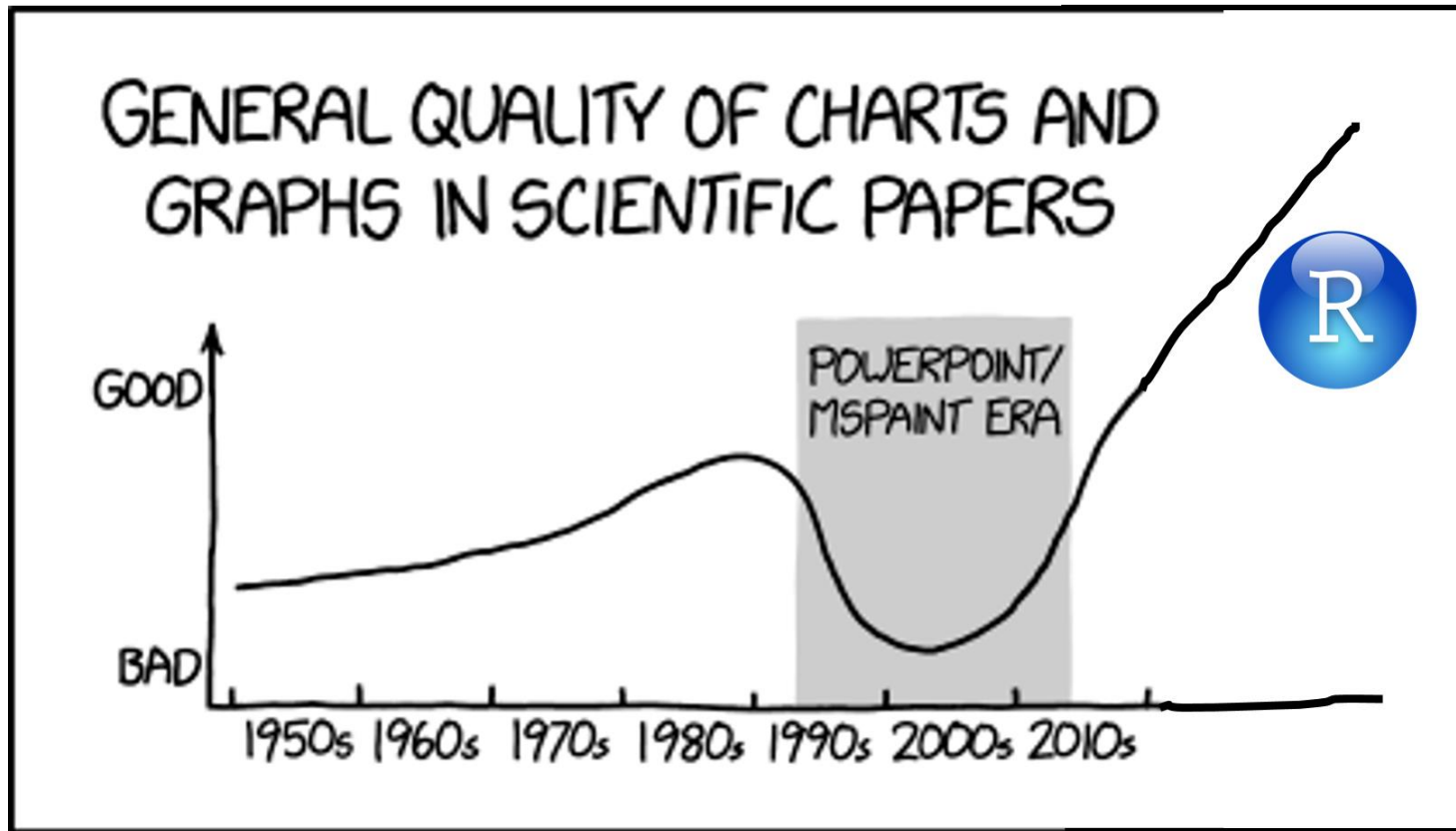
- Bruk **by** = for å spesifisere nøkkel variabel til å koble på

```
by = c("aar" = "year")
```

```
kobletdata <- left_join(datanavn1, datanavn2, by = variabelnavn)
```

- Flere variabler kan brukes for å koble på (som en vektor)

Plotting



Plotting med ggplot()

- **aes** : aesthetics, hvilke variabler
- **geom_** : hva slags figur
- **stat** : hva slags statistisk aggregat å presentere

Table 18-1 A Selection of Geoms and Associated Default Stats

<i>Geom</i>	<i>Description</i>	<i>Default Stat</i>
<code>geom_bar()</code>	Bar chart	<code>stat_bin()</code>
<code>geom_point()</code>	Scatterplot	<code>stat_identity()</code>
<code>geom_line()</code>	Line diagram, connecting observations in order by x-value	<code>stat_identity()</code>
<code>geom_boxplot</code>	Box-and-whisker plot	<code>stat_boxplot()</code>
<code>geom_path</code>	Line diagram, connecting observations in original order	<code>stat_identity()</code>
<code>geom_smooth</code>	Add a smoothed conditioned mean	<code>stat_smooth()</code>
<code>geom_histogram</code>	An alias for <code>geom_bar()</code> and <code>stat_bin()</code>	<code>stat_bin()</code>



Søylediagram

```
ggplot(aes(variabelnavn)) +  
  geom_bar()
```

Bruke + for å legge til figurtype

Spesifisere variabelen

Spesifisere søylediagram

```
ggplot(aes=c(x = variabelnavn1, y = variabelnavn2)) +  
  geom_bar(stat="identity")
```

Spesifisere x og y variablene

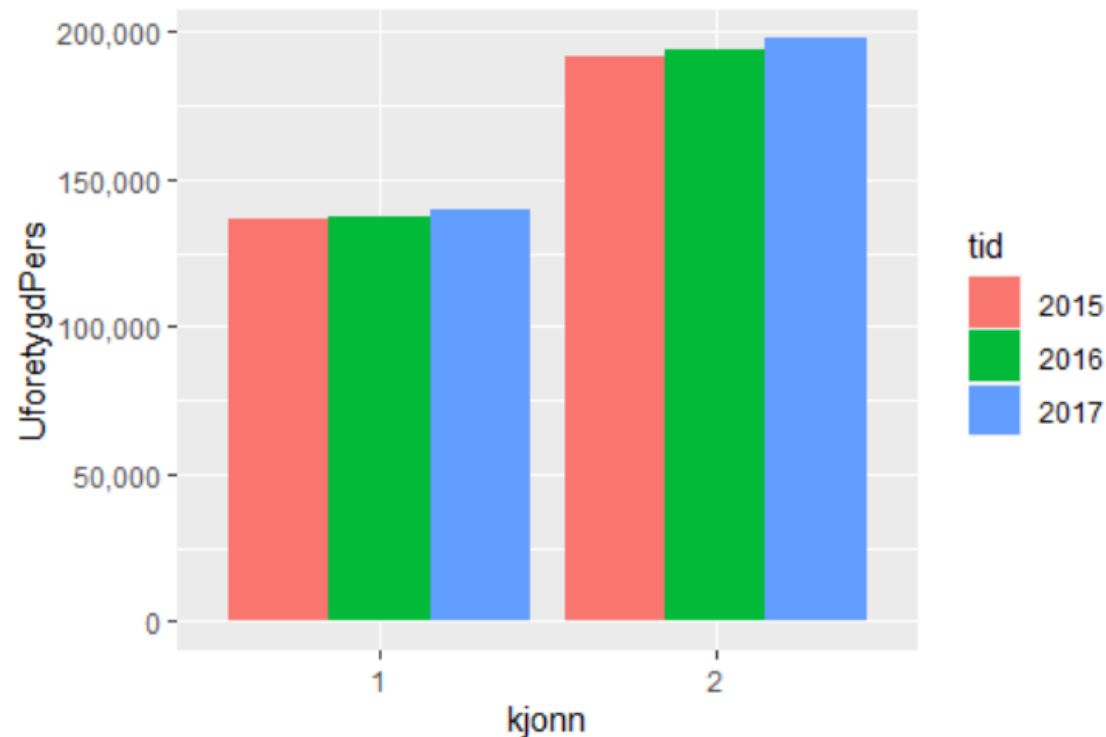
Spesifisere å bruke verdi



Statistisk sentralbyrå
Statistics Norway

Søylediagram

- Bruk fill() i aes for å spesifisere en variabel for farge
- Kombinere med filter først



Punktdiagram

- Sammenlign to numeriske variabler

```
ggplot(aes(x = variabelnavn1, y = variabelnavn2)) +  
geom_point()
```


- Legg til regresjonslinje med

```
geom_smooth(method = "lm")
```

- Farge punktene etter gruppe

```
geom_point(aes(color = variabelnavn))
```

Lagre figur

- Klikk  Export ▾
- Eller lagre på «arbeidsområde» (`getwd()`)

```
png(file = "figurnavn.png")  
ggplot(aes(variabelnavn)) +  
  geom_bar()  
dev.off()
```

Spesifisere filnavn

Lage plot

Spesifisere vi er ferdig

Eksportere en tabell til excel

```
library(openxlsx)  
write.xlsx(datanavn, file = "datafilenavn.xlsx")
```


Øvelse 4

- Gå inn til samme møte/chatrommet som tidligere.
- Øvelsene til oppgavesett 1 er på fil: **øvelser_4.R**



Oppsummering

- Husk library()
- Les inn filer: read_csv2() read_sas()
- Ny variabel: mutate()
- Velg noen linje: filter()
- Aggregere/oppsummere: summarise()
- Figur: ggplot(), aes(), geom_...()
- <https://wiki.ssb.no/display/s880/For+R+brukere>
- Yammer: R i SSB
- Google

