

Oppgave 4

```
rm(list=ls())

library(rvest)
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5      v purrr 0.3.4
## v tibble 3.1.4      v dplyr 1.0.7
## v tidyr 1.1.3       v stringr 1.4.0
## v readr 2.0.1       v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter()    masks stats::filter()
## x readr::guess_encoding() masks rvest::guess_encoding()
## x dplyr::lag()       masks stats::lag()

library(rlist)

## Warning: package 'rlist' was built under R version 4.1.2

library(purrr)
library(zoo)

## Warning: package 'zoo' was built under R version 4.1.2

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
```

Først hentes alle kalenderene gjennom hver sin URL og lagres som en liste.

```
urls <- list("https://timeplan.uit.no/emne_timeplan.php?sem=22v&module%5B%5D=SOK-1005-1&week=1-20&View=",
             "https://timeplan.uit.no/emne_timeplan.php?sem=22v&module%5B%5D=SOK-1016-1&week=1-20&View=",
             "https://timeplan.uit.no/emne_timeplan.php?sem=22v&module%5B%5D=SOK-1006-1&week=1-20&View=")
```

Så lages en funksjon som henter informasjonen fra UiT sin timeplan, omstrukturerer den og velger ut de viktige kolonnene. Merk at på dager det er flere timer vil kun første time få verdier for dato, derfor måtte jeg bruke pakken zoo og legge inn verdien fra raden over disse datoene.

```

scraper <- function(url) {
  page <- read_html(url)

  table <- html_nodes(page, 'table')
  table <- html_table(table, fill=TRUE)

  dframe <- list.stack(table)

  colnames(dframe) <- dframe[1,]

  dframe <- dframe %>% filter(!Dato=="Dato")

  dframe$Dato[dframe$Dato == ""] <- NA
  dframe$Dato <- na.locf(dframe$Dato)

  dframe <- dframe %>% separate(Dato,
                                into = c("Dag", "Dato"),
                                sep = "(?<=[A-Za-z])(?=[0-9])")

  dframe$Dato <- as.Date(dframe$Dato, format="%d.%m.%Y")
  dframe$Uke <- strptime(dframe$Dato, format = "%V")

  dframe <- dframe %>%
    arrange(Dato)

  dframe <- dframe %>% select(Dag,Dato,Uke,Tid,Rom,Emnekode)

  return(dframe)
}

```

Til slutt så lagres alle observasjonene i en tabell og den sorteres etter dato.

```
calendar <- map(urls, scraper)
```

```
## Column 1 ['Uke 3'] of item 2 is missing in item 1. Use fill=TRUE to fill with NA (NULL for list column)
## Column 1 ['Uke 3'] of item 2 is missing in item 1. Use fill=TRUE to fill with NA (NULL for list column)
```

```
## Column 1 ['Uke 4'] of item 2 is missing in item 1. Use fill=TRUE to fill with NA (NULL for list column)
```

```
calendar <- bind_rows(calendar)
```

```
calendar <- calendar %>%
  arrange(Dato, Tid)
calendar
```

```
##      Dag      Dato Uke      Tid      Rom  Emnekode
## 1  Onsdag 2022-01-12 02 14:15-16:00 VIRTUELT _Digital undervisning SOK-1016-1
## 2  Torsdag 2022-01-13 02 14:15-16:00 VIRTUELT _Digital undervisning SOK-1005-1
## 3  Tirsdag 2022-01-18 03 10:15-12:00 VIRTUELT _Digital undervisning SOK-1016-1
## 4  Tirsdag 2022-01-18 03 12:15-13:00      HHT 01.202AUD SOK-1006-1
```

## 5	Onsdag	2022-01-19	03	12:15-16:00	VIRTUELT	_Digital undervisning	SOK-1006-1
## 6	Torsdag	2022-01-20	03	09:15-14:00	VIRTUELT	_Digital undervisning	SOK-1006-1
## 7	Torsdag	2022-01-20	03	14:15-16:00	VIRTUELT	_Digital undervisning	SOK-1005-1
## 8	Tirsdag	2022-01-25	04	14:15-17:00		HARSTAD_HG5 AUD_GRØNT	SOK-1006-1
## 9	Torsdag	2022-01-27	04	12:15-14:00	VIRTUELT	_Digital undervisning	SOK-1016-1
## 10	Torsdag	2022-01-27	04	14:15-16:00	VIRTUELT	_Digital undervisning	SOK-1005-1
## 11	Torsdag	2022-02-03	05	14:15-16:00		HHT 01.202AUD	SOK-1005-1
## 12	Torsdag	2022-02-10	06	14:15-16:00		HHT 01.202AUD	SOK-1005-1
## 13	Fredag	2022-02-11	06	12:15-14:00		TEO-H1 1.836-AUD2	SOK-1016-1
## 14	Tirsdag	2022-02-15	07	12:15-14:00		TEO-H1 1.836-AUD2	SOK-1016-1
## 15	Onsdag	2022-02-16	07	12:15-16:00		MH U6.SAUD	SOK-1006-1
## 16	Torsdag	2022-02-17	07	08:15-10:00		TEO-H6 6.303-AUD3	SOK-1006-1
## 17	Torsdag	2022-02-17	07	12:15-14:00		TEO-H6 6.303-AUD3	SOK-1006-1
## 18	Torsdag	2022-02-17	07	14:15-16:00		HHT 01.202AUD	SOK-1005-1
## 19	Torsdag	2022-02-24	08	11:15-14:00		TEO-H1 1.820-AUD1	SOK-1006-1
## 20	Torsdag	2022-02-24	08	14:15-16:00		HHT 01.202AUD	SOK-1005-1
## 21	Fredag	2022-02-25	08	12:15-14:00		TEO-H1 1.836-AUD2	SOK-1016-1
## 22	Mandag	2022-02-28	09	10:15-12:00		HHT 02.217	SOK-1006-1
## 23	Mandag	2022-02-28	09	12:15-14:00		TEO-H1 1.836-AUD2	SOK-1016-1
## 24	Tirsdag	2022-03-01	09	12:15-14:00		HHT 02.218	SOK-1016-1
## 25	Onsdag	2022-03-02	09	10:15-12:00		HHT 02.217	SOK-1006-1
## 26	Onsdag	2022-03-02	09	14:15-16:00		HHT 02.218	SOK-1016-1
## 27	Torsdag	2022-03-03	09	14:15-16:00		HHT 01.202AUD	SOK-1005-1
## 28	Torsdag	2022-03-10	10	14:15-16:00		HHT 01.202AUD	SOK-1005-1
## 29	Mandag	2022-03-14	11	10:15-12:00		HHT 02.217	SOK-1006-1
## 30	Tirsdag	2022-03-15	11	10:15-12:00		TEO-H1 1.836-AUD2	SOK-1016-1
## 31	Onsdag	2022-03-16	11	10:15-12:00		HHT 02.217	SOK-1006-1
## 32	Onsdag	2022-03-16	11	12:15-16:00		TEO-H6 6.303-AUD3	SOK-1006-1
## 33	Onsdag	2022-03-16	11	14:15-16:00		HHT 02.218	SOK-1016-1
## 34	Torsdag	2022-03-17	11	14:15-16:00		HHT 01.202AUD	SOK-1005-1
## 35	Fredag	2022-03-18	11	12:15-14:00		HHT 02.218	SOK-1016-1
## 36	Mandag	2022-03-21	12	10:15-12:00		HHT 02.217	SOK-1006-1
## 37	Tirsdag	2022-03-22	12	10:15-12:00		TEO-H1 1.836-AUD2	SOK-1016-1
## 38	Onsdag	2022-03-23	12	10:15-12:00		HHT 02.217	SOK-1006-1
## 39	Onsdag	2022-03-23	12	14:15-16:00		HHT 02.218	SOK-1016-1
## 40	Torsdag	2022-03-24	12	14:15-16:00		HHT 01.202AUD	SOK-1005-1
## 41	Fredag	2022-03-25	12	10:15-12:00		HHT 02.218	SOK-1016-1
## 42	Mandag	2022-03-28	13	10:15-12:00		HHT 02.217	SOK-1006-1
## 43	Tirsdag	2022-03-29	13	16:00-18:00		TEO-H6 6.303-AUD3	SOK-1006-1
## 44	Onsdag	2022-03-30	13	09:15-11:00		FARM F1.101	SOK-1006-1
## 45	Onsdag	2022-03-30	13	10:15-12:00		HHT 02.217	SOK-1006-1
## 46	Onsdag	2022-03-30	13	12:15-14:00		FARM F1.101	SOK-1006-1
## 47	Onsdag	2022-03-30	13	14:15-16:00		HHT 02.218	SOK-1016-1
## 48	Torsdag	2022-03-31	13	14:15-16:00		HHT 01.202AUD	SOK-1005-1
## 49	Fredag	2022-04-01	13	12:15-14:00		HHT 02.218	SOK-1016-1
## 50	Mandag	2022-04-04	14	10:15-12:00		HHT 02.217	SOK-1006-1
## 51	Tirsdag	2022-04-05	14	12:15-14:00		TEO-H1 1.836-AUD2	SOK-1016-1
## 52	Onsdag	2022-04-06	14	10:15-12:00		HHT 02.217	SOK-1006-1
## 53	Onsdag	2022-04-06	14	14:15-16:00		HHT 02.218	SOK-1016-1
## 54	Torsdag	2022-04-07	14	11:15-14:00		TEO-H1 1.820-AUD1	SOK-1006-1
## 55	Torsdag	2022-04-07	14	14:15-16:00		HHT 01.202AUD	SOK-1005-1
## 56	Fredag	2022-04-08	14	12:15-14:00		HHT 02.218	SOK-1016-1
## 57	Tirsdag	2022-04-19	16	10:15-12:00		TEO-H1 1.836-AUD2	SOK-1016-1
## 58	Tirsdag	2022-04-19	16	12:15-14:00		HHT 02.217	SOK-1006-1

## 59	Onsdag	2022-04-20	16	10:15-12:00	HHT 02.217	SOK-1006-1
## 60	Onsdag	2022-04-20	16	14:15-16:00	HHT 02.218	SOK-1016-1
## 61	Torsdag	2022-04-21	16	12:15-14:00	HHT 02.218	SOK-1016-1
## 62	Torsdag	2022-04-21	16	14:15-16:00	HHT 01.202AUD	SOK-1005-1
## 63	Fredag	2022-04-22	16	09:15-12:00	TEO-H1 1.836-AUD2	SOK-1006-1
## 64	Mandag	2022-04-25	17	10:15-12:00	HHT 02.217	SOK-1006-1
## 65	Tirsdag	2022-04-26	17	10:15-12:00	TEO-H1 1.836-AUD2	SOK-1016-1
## 66	Onsdag	2022-04-27	17	10:15-12:00	HHT 02.217	SOK-1006-1
## 67	Onsdag	2022-04-27	17	12:15-16:00	TEO-H1 1.836-AUD2	SOK-1006-1
## 68	Onsdag	2022-04-27	17	14:15-16:00	HHT 02.218	SOK-1016-1
## 69	Torsdag	2022-04-28	17	09:15-14:00	TEO-H1 1.820-AUD1	SOK-1006-1
## 70	Torsdag	2022-04-28	17	12:15-14:00	HHT 02.218	SOK-1016-1
## 71	Torsdag	2022-04-28	17	14:15-16:00	HHT 01.202AUD	SOK-1005-1
## 72	Mandag	2022-05-02	18	10:15-12:00	HHT 02.217	SOK-1006-1
## 73	Tirsdag	2022-05-03	18	10:15-12:00	TEO-H1 1.836-AUD2	SOK-1016-1
## 74	Onsdag	2022-05-04	18	10:15-12:00	HHT 02.217	SOK-1006-1
## 75	Onsdag	2022-05-04	18	14:15-16:00	HHT 02.218	SOK-1016-1
## 76	Torsdag	2022-05-05	18	12:15-14:00	HHT 02.218	SOK-1016-1
## 77	Torsdag	2022-05-05	18	14:15-16:00	HHT 01.202AUD	SOK-1005-1
## 78	Mandag	2022-05-09	19	10:15-12:00	HHT 02.217	SOK-1006-1
## 79	Onsdag	2022-05-11	19	08:15-10:00	TEO-H1 1.836-AUD2	SOK-1016-1
## 80	Onsdag	2022-05-11	19	10:15-12:00	HHT 02.217	SOK-1006-1
## 81	Onsdag	2022-05-11	19	14:15-16:00	HHT 02.218	SOK-1016-1
## 82	Torsdag	2022-05-12	19	12:15-14:00	HHT 02.218	SOK-1016-1
## 83	Fredag	2022-05-13	19	13:00-13:00		SOK-1005-
## 84	Mandag	2022-05-16	20	10:15-12:00	HHT 02.217	SOK-1006-1
## 85	Onsdag	2022-05-18	20	08:15-10:00	TEO-H1 1.836-AUD2	SOK-1016-1
## 86	Onsdag	2022-05-18	20	10:15-12:00	HHT 02.217	SOK-1006-1
## 87	Onsdag	2022-05-18	20	14:15-16:00	HHT 02.218	SOK-1016-1
## 88	Torsdag	2022-05-19	20	12:15-14:00	HHT 02.218	SOK-1016-1