Consortium for
Statistics in
Disease Surveillance

UNIVERSITY
OF OSLO

Introducing `csverse`

CSIDS approach to Automated Public Health Surveillance

Chi Zhang, PhD chi.zhang@medisin.uio.no

CSIDS, Oslo Center for Biostatistics and Epidemiology @ University of Oslo
ex-NIPH (Norwegian Institute of Public Health)

2023-12-07 Northern European Symposium on Automated Surveillance

About this talk

Situational reports

The old way
&
The CSIDS way

CSIDS::csverse

The R ecosystem for automated PH surveillance

Ways forward

“How to prevent the next pandemic”

Disclaimers

The opinions are my own, and do not reflect the views of my employer.

Regulations and policies are constantly changing. Please check the official documents for each data source for most updated information.

Situational reports

Daily situational reports

Nation, 11 counties, 356 municipalities

Situational reports on **covid cases, hospitalisation, vaccine** among other outcomes of interest

Used by FHI leadership and Ministry of Health

Historically (pre 2020.12) made **manually** for **selected locations per request**

Need to deliver **in the early morning: overtime 6–8am**

Expensive, time consuming, prone to error

How can we do better?

COVID-19

Dagsrapport – nasjonalt og per fylke

onsdag 20. oktober 2021

FHI Folkehelseinstituttet

Demo

Dagens foreløpige (u.off) tall. Rapporten er generert kl. . Der annet ikke er oppgitt er figurene basert på prøvedato i MSIS. Tallene er midlertidige og kan bli endret.

Status oppdatering

Totalt nye tilfeller ble registrert siste døgn. De to foregående dagene ble det registrert henholdsvis tilfeller.

Til sammenligning ble det for en uke siden, den 06.10.2021 rapportert registrerte tilfeller siste døgn.

Antall meldte basert på prøvedato så langt uke 41 er mot på samme tid sist uke (uke 40). Antall meldte basert på registrert dato så langt denne uka (uke 41) er , mot på samme tid sist uke (40).

Totalt er personer vaksinert mot covid-19 i Norge, av disse er personer vaksinert med både 1 og 2. dose med koronavaksine av hele befolkningen er vaksinert med minst en dose og vaksinert med to doser med koronavaksine. Blant personer 18 år og eldre er . % vaksinert med minst en dose med koronavaksine, og av disse er % vaksinert med to doser med koronavaksine. Blant personer 45 år og eldre er % vaksinert med minst en dose, og % av personer 65 år og eldre er vaksinert med minst en dose med koronavaksine. Data er hentet fra BeredtC19, SYSVAK, per 12.10.2021.

Open source transition

Why should I switch?

Consider the following aspects:

Cost

Ease of use (e.g well documented)

Efficiency (e.g. automation)

Collaboration and teamwork

Reproducibility

Research and new method adoption

From

To

Daily situational reports: automated

Nation, 11 counties, 356 municipalities

Everyday, before **7am**

Delivered to N folder so that everyone with access can **view / modify**

35 pages, 17 tables, 21 figures (e.g. national report)

Consistent style with NIPH documents

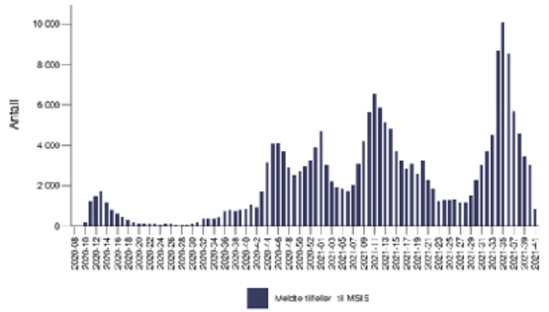
- Agder_fylke_dagsrapport_covid19_2021-10-20.docx
- Innlandet_fylke_dagsrapport_covid19_2021-10-20.docx
- Møre_og_Romsdal_fylke_dagsrapport_covid19_2021-10-20.docx
- Nordland_fylke_dagsrapport_covid19_2021-10-20.docx
- Oslo_fylke_dagsrapport_covid19_2021-10-20.docx
- Rogaland_fylke_dagsrapport_covid19_2021-10-20.docx
- Troms_og_Finnmark_fylke_dagsrapport_covid19_2021-10-20.docx
- Trøndelag_fylke_dagsrapport_covid19_2021-10-20.docx
- Vestfold_og_Telemark_fylke_dagsrapport_covid19_2021-10-20.docx
- Vestland_fylke_dagsrapport_covid19_2021-10-20.docx
- Viken_fylke_dagsrapport_covid19_2021-10-20.docx

Tabell 1 Covid-19 status og utvikling, uke 2021-37 til uke 2021-41.

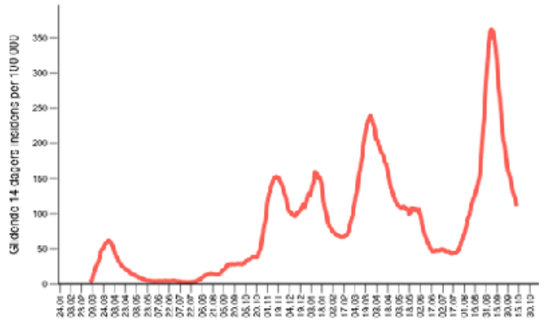
Overvåkingssystem/ Indikatorer	2021-37	2021-38	2021-39	2021-40	2021-41	Totalt av de siste 5 ukene
Meldte tilfeller til MSIS (prøvedato)						
Meldte tilfeller til MSIS (registrertdato)						
Antall personer testet for SARS-CoV-2 (PCR)						
Nye covid-19 positive pasienter innlagt i sykehus (alle årsaker)						
Nye pasienter innlagt i sykehus med covid-19 som hoved-årsak						
Nye pasienter med bekreftet covid-19 innlagt i intensiv-avdeling						
Covid-19-assosierte dødsfall						

*Visualiseringen på nettsidene våre oppdateres ca kl 13.00 hver dag.

Merk: Denne tabellen og visualiseringen på www.fhi.no viser antall personer testet med PCR, vi jobber med å inkludere antigen hurtigtester. Se ukesrapport for framstilling av antall testede med PCR og hurtigtester samlet. Totalt har [redacted] ersoner blitt diagnostisert i Norge frem til kl 24.00, 12.10.2021, og tilsammen [redacted] personer er registrert testet for covid-19 (per 12.10.2021).



Figur 1. Antall tilfeller av covid-19 per uke basert på prøvedato gjennom hele pandemiperioden, Norge.



Figur 2. Glidende 14-dagers insidens per 100 000 innbyggere per dag basert på prøvedato gjennom hele pandemiperioden, Norge.

Automation saves time and money

Among other good things

National + 11 counties + 356 municipalities = 368 reports
From Secure zone (data extraction) to reports = **40**min (8 CPU in parallel)

One year deliverables	Manual report	CSIDS / csverse
Number of reports	5 * 365 = 1 825	368 * 365 = 134 320
Overtime hours	700	0
Number of people needed	4	0
Overtime costs	700 000 kr	0

700 000 kr is approximately 67 400 Euro, 70 245 USD (2022.06.15)

Overtime estimated by one employee work from 6 to 8am for one year (2h per day)
One employee is allowed 200h overtime per year

CSIDS::csverse

CSIDS

Consortium for Statistics in Disease Surveillance

2022.11 – **CSIDS** (reads see-sids) was founded, collaboration between Norwegian Institute of Public Health, University of Oslo

Promote **collaborative development** of open source R packages for automated public health **surveillance and reporting**

Standardised, reproducible, open and free

Anyone is welcome to use our R packages, and welcome to contribute

www.csids.no

csverse

CSIDS R package ecosystem

Public health surveillance algorithms

csalert, attrib, nowcast

Analysis planning and file organization

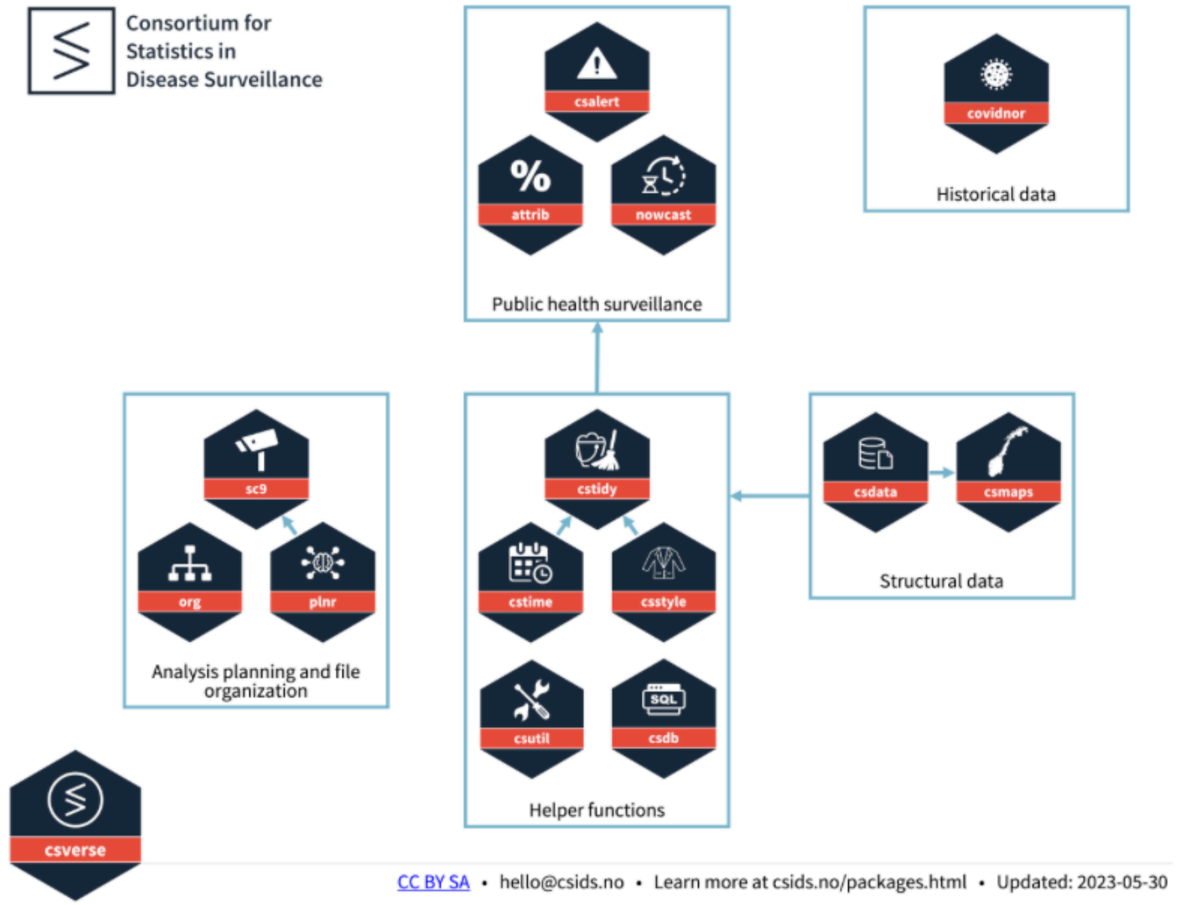
sc9, org, plnr

Historical and structured data

csdata, csmaps, covidnor (now respiranor)

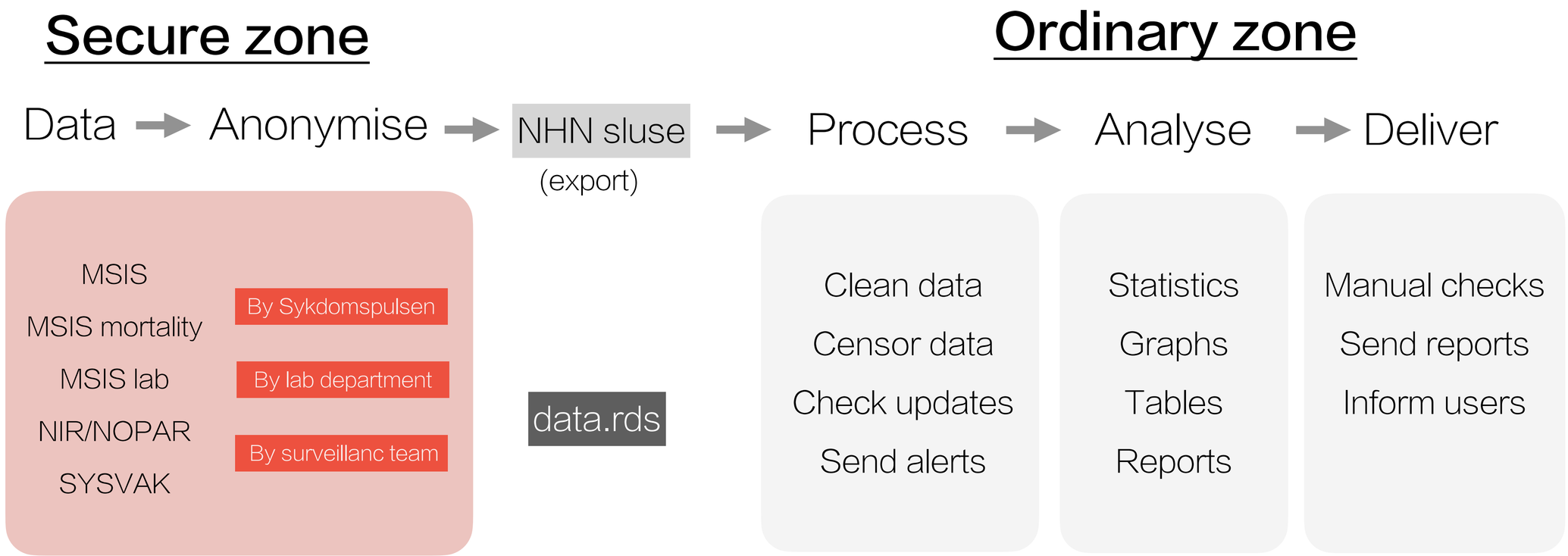
Helper functions

cstidy, cstime, csstyle, csdb, csutil



Packages are under active development, please check each pkg documentation for latest update

From data to report



Sykdomspulsen: CSIDS until 2022.11
MSIS: Norwegian Surveillance System for Communicable Diseases
NIR: Norwegian Intensive Care Registry
SYSVAK: Norwegian Immunisation Registry
NOPAR: Norwegian Pandemic Registry

cstidy cstime

Produce standardized formatted data

We use **data.table** instead of **dplyr** for efficient data manipulation

CSIDS style guide for **names** of variable on time, location, age groups, sex etc

Fixed fields (granularity_time, granularity_geo , …) available for all datasets

Task specific fields: used for computation

https://www.csids.no/cstidy/articles/csfmt_rts_data_v1.html

```
d <- cstidy::generate_test_data()[1:5]
cstidy::set_csfmt_rts_data_v1(d)

# Looking at the dataset
d[]
#>    granularity_time granularity_geo country_iso3 location_code border  age  sex
#> 1:      isoyearweek          county          nor  county_nor42     NA <NA> <NA>
#> 2:      isoyearweek          county          nor  county_nor32     NA <NA> <NA>
#> 3:      isoyearweek          county          nor  county_nor33     NA <NA> <NA>
#> 4:      isoyearweek          county          nor  county_nor56     NA <NA> <NA>
#> 5:      isoyearweek          county          nor  county_nor34     NA <NA> <NA>
#>    isoyear isoweek isoyearweek    season seasonweek calyear calmonth
#> 1:    2022       3     2022-03 2021/2022         26      NA       NA
#> 2:    2022       3     2022-03 2021/2022         26      NA       NA
#> 3:    2022       3     2022-03 2021/2022         26      NA       NA
#> 4:    2022       3     2022-03 2021/2022         26      NA       NA
#> 5:    2022       3     2022-03 2021/2022         26      NA       NA
#>    calyearmonth       date deaths_n
#> 1:         <NA> 2022-01-23        6
#> 2:         <NA> 2022-01-23        4
#> 3:         <NA> 2022-01-23        6
#> 4:         <NA> 2022-01-23        3
#> 5:         <NA> 2022-01-23        6
```

- **covid19_testevents/_n**: Number of covid-19 test events (i.e. a person getting tested within a 7 day period).
- **covid19_testevents_pos/_pr1**: Proportion (0-1) of covid-19 test events that were positive.
- **covid19_testevents_pos/_pr100**: Percentage (0-100) of covid-19 test events that were positive.
- **covid19_testevents_pos/_sum0_13/_pr100**: Percentage (0-100) of covid-19 test events that were positive over the last 14 days.
- **covid19_testevents_pos/_daymean0_13/_pr100**: For each of the last 14 days, calculate the percentage (0-100) of covid-19 test events that were positive, and then take the mean of these 14 values.
- **covid19_testevents_pos/_isoweekmean0_13/_pr100**: For each of the last 7 day periods (0-6 days, 7-13 days), calculate the percentage (0-100) of covid-19 test events that were positive, and then take the mean of these 2 values.

csalert

Analyse trend, produce predictions

Consortium for Statistics in Disease Surveillance

csalert 2023.6.26 Articles Reference News

Function reference

All functions

`add_holiday_effect()`
Holiday effect ----

`prediction_interval()`
Prediction thresholds

`prediction_interval(<glm>)`
Prediction thresholds

`short_term_trend()`
Determine the short term trend of a timeseries

`signal_detection_hlm()`
Determine the short term trend of a timeseries

`simulate_baseline_data()`
Simulate baseline data ---- Simulation of baseline data.

`simulate_seasonal_outbreak_data()`
Simulate seasonal outbreaks ----

`simulate_spike_outbreak_data()`
Simulate spiked outbreaks ----

Determine the short term trend of a timeseries

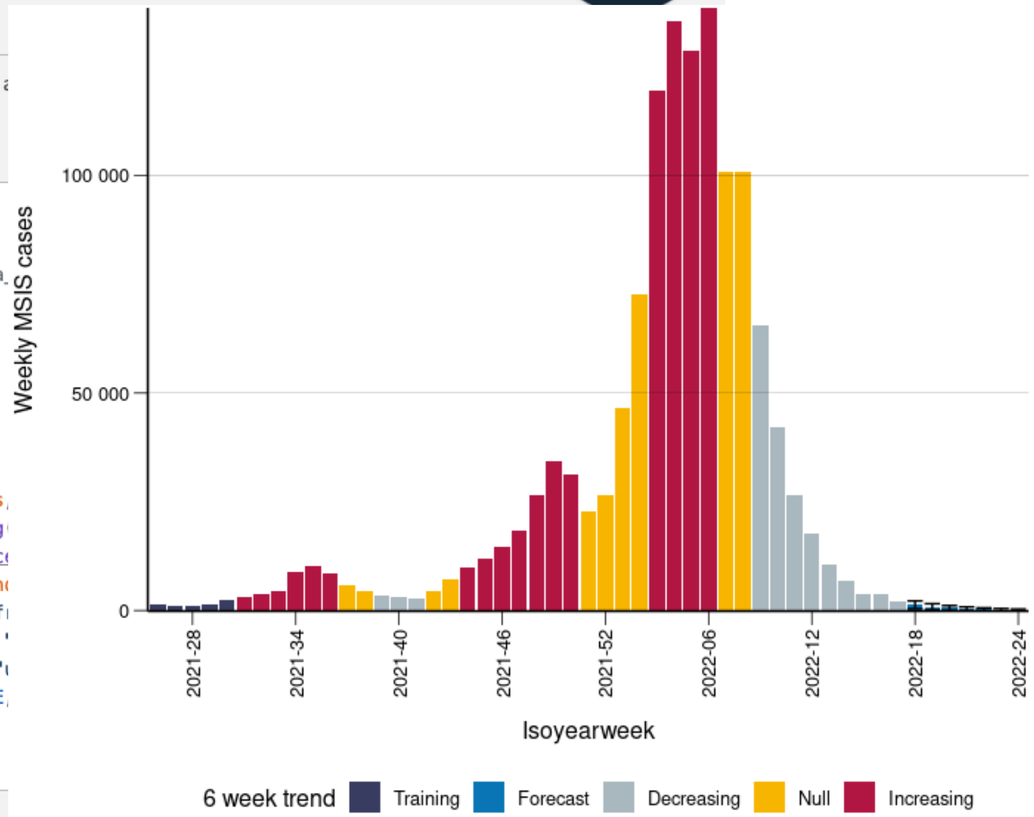
Source: R/short_term_trend.R

The method is based upon a published a

Usage

```
short_term_trend(x, ...)

# S3 method for csfmt_rts_data
short_term_trend(
  x,
  numerator,
  denominator = NULL,
  prX = 100,
  trend_dates = 42,
  remove_last_dates = 0,
  forecast_dates = trend_dates,
  trend_isoyearweeks = ceiling
  remove_last_isoyearweeks = c
  forecast_isoyearweeks = tren
  numerator_naming_prefix = "f
  denominator_naming_prefix = '
  statistics_naming_prefix = "
  remove_training_data = FALSE,
  ...
)
```



csmaps

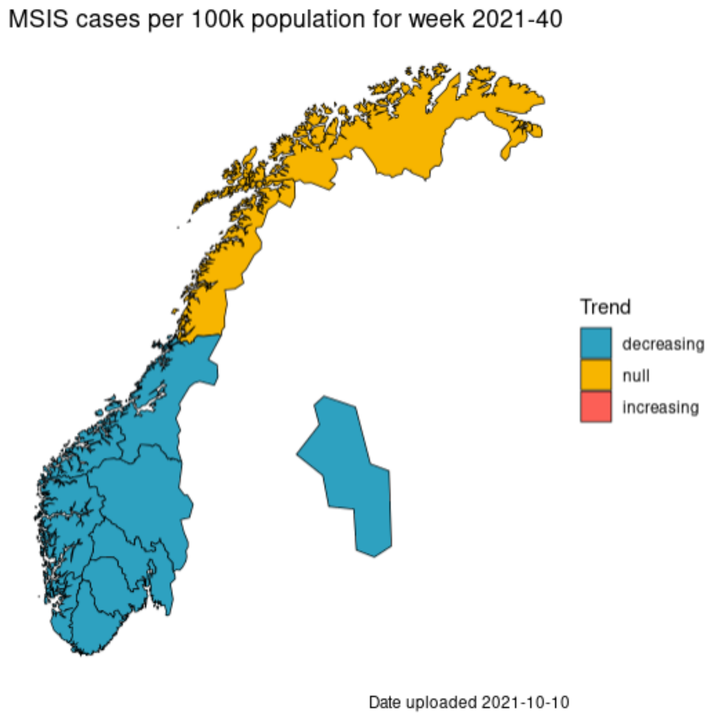
Create maps with rich information

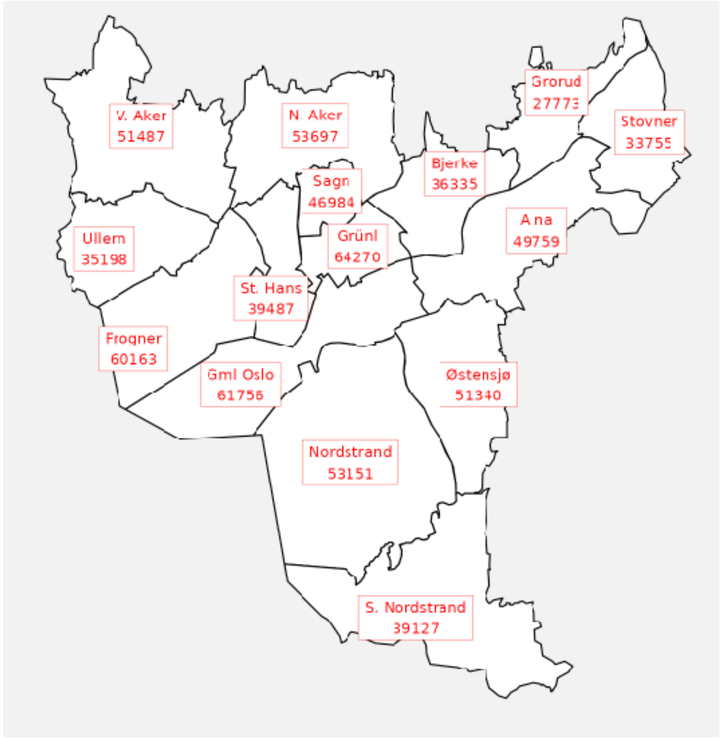
Norway maps for **county, municipality** levels and **Oslo**

```
pd <- copy(csmaps::nor_municip_map_b2024_default_dt)
q <- ggplot()
q <- q + geom_polygon(
  data = pd,
  aes(
    x = long,
    y = lat,
    group = group
  ),
  color="black",
  fill="white",
  linewidth = 0.2
)
q <- q + theme_void()
q <- q + coord_quickmap()
q <- q + labs(title = "Default layout")
q
```

csmaps

Create maps with rich information





plnr

Structured analysis planning

Structured way of planning analysis combinations for large dataset
(e.g. 400 **locations** * 10 yr **daily** data * 10 **age groups** * 50 **codes**)

- pull once from DB
- develop **analysis code for one subset**
- repeat for all combinations. Can be paralleled!

Plan

```
location_code = c('norge', 'county03', 'county15')

age_group = c('00_04', '05_14')

tag_outcome = c('covid19_n_cases', 'influenza_n_cases')
```

```
# 2. add argset
# check location codes
location_codes <- p$get_data()$covid19_cases$location_code %>%
  unique() %>%
  print()

p$add_argset_from_list(
  plnr::expand_list(
    location_code = location_codes,
    granularity_time = "isoweek"
  )
)
# Examine the argsets that are available
p$get_argsets_as_dt()
```

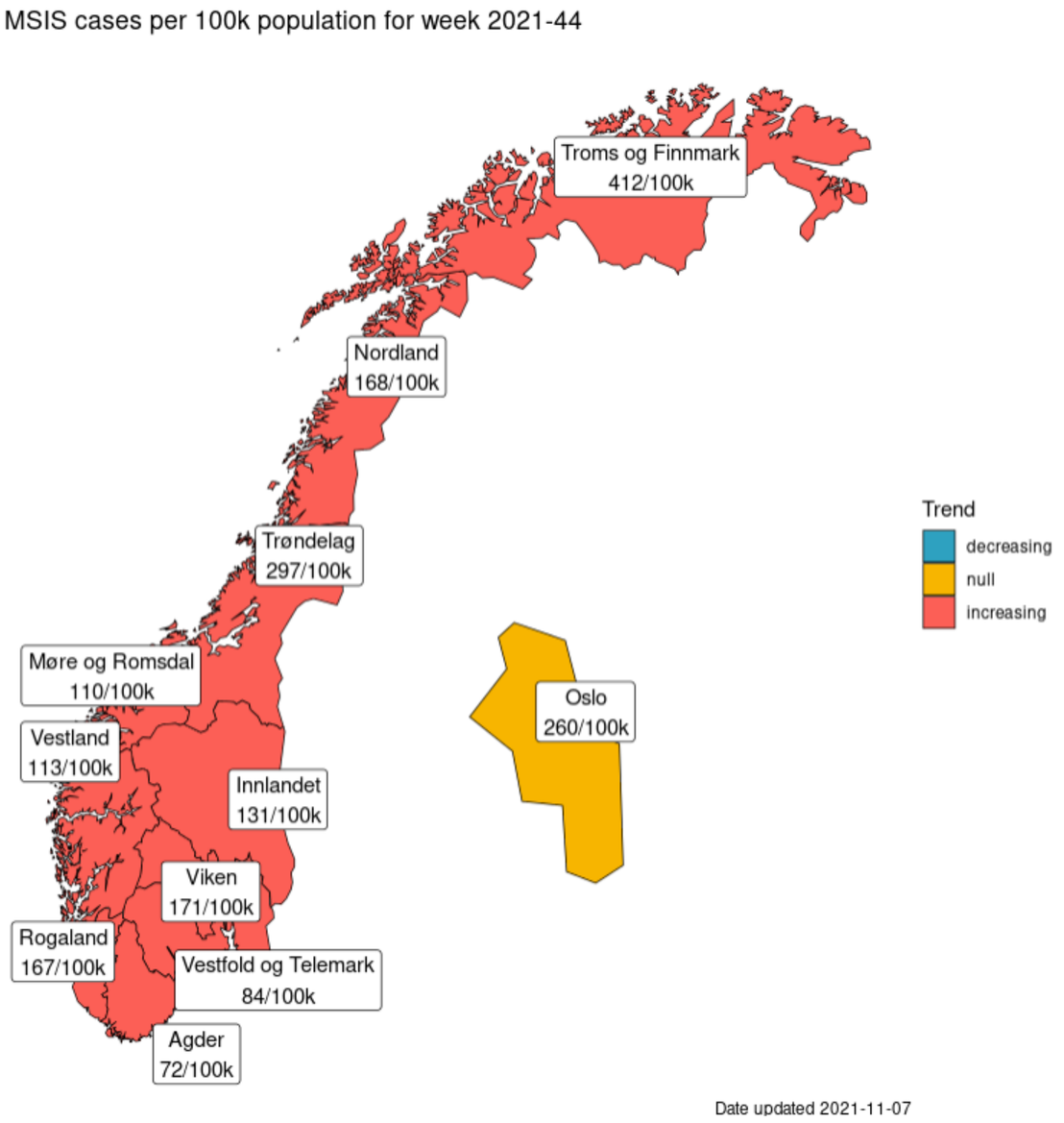
Expanded list of combinations

```
Plan1: location_code = 'norge', age_group = '00_04', tag_outcome = 'covid19_n_cases'
Plan2: location_code = 'county03', age_group = '00_04', tag_outcome = 'covid19_n_cases'
...
```

Visualize trends

Trend for 11 counties

```
> d_msis_this_isoyearweek
    location_code       date isoyearweek covid19_cases_testdate_pr100000 c
 1:      county03 2021-11-07     2021-44                        260.25452
 2:      county11 2021-11-07     2021-44                        167.41083
 3:      county15 2021-11-07     2021-44                        109.58636
 4:      county18 2021-11-07     2021-44                        167.67563
 5:      county30 2021-11-07     2021-44                        170.63457
 6:      county34 2021-11-07     2021-44                        130.86780
 7:      county38 2021-11-07     2021-44                         84.14675
 8:      county42 2021-11-07     2021-44                         71.55739
 9:      county46 2021-11-07     2021-44                        112.86417
10:      county50 2021-11-07     2021-44                        297.16168
11:      county54 2021-11-07     2021-44                        411.60766
```



Automated reports

Requires

- `style.docx`
- `report.Rmd`

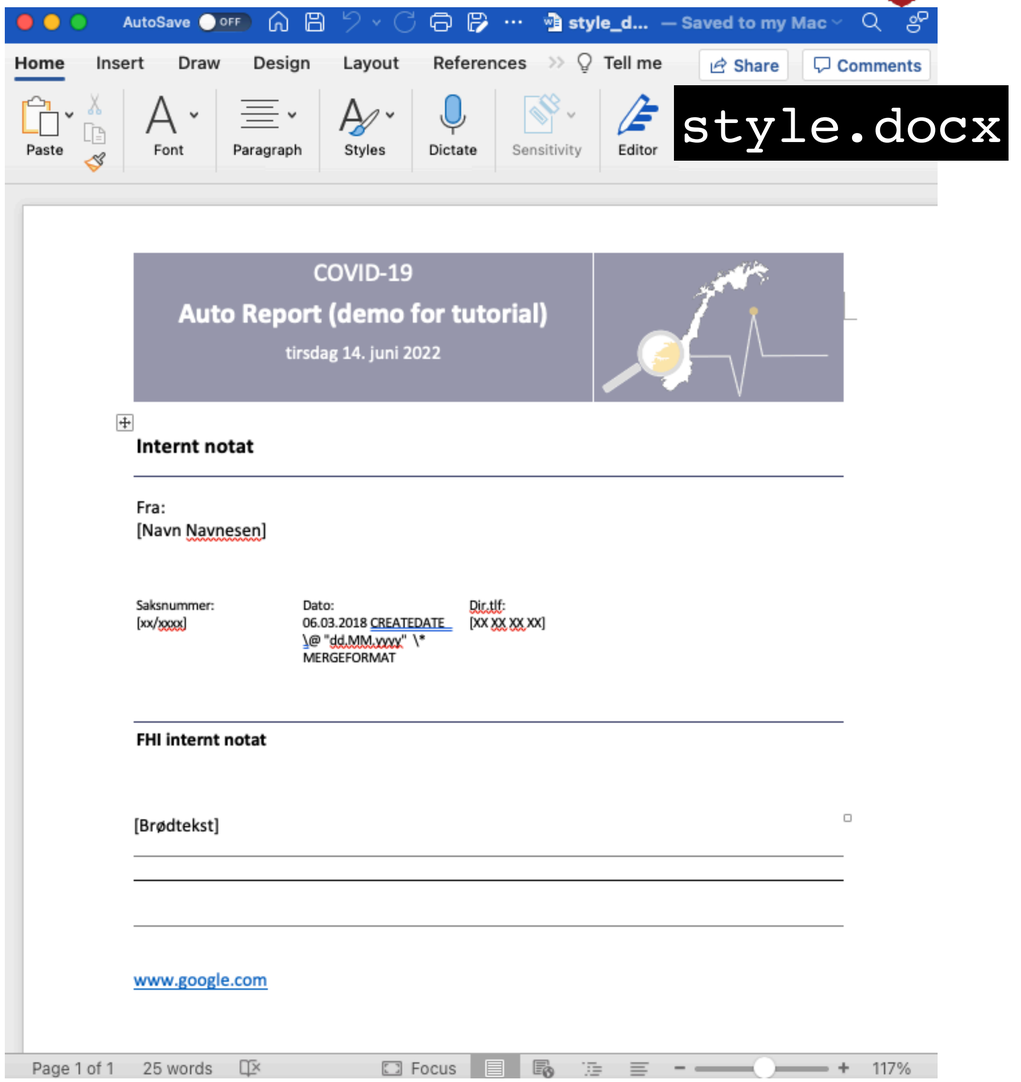
Input

- `report.docx`

Output

- `script.R`

Automate



Automated reports

report.Rmd

```
---
title: "Test Report"
output:
    reference_docx: style_demo.docx

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

Introduction and disclaimers

This is a demonstration for how to make automated reports with Rmarkdown and MS Word.

The dates, numbers, graphs and tables are for educational purposes only. The text is made to teach R, not to be used in other circumstances.

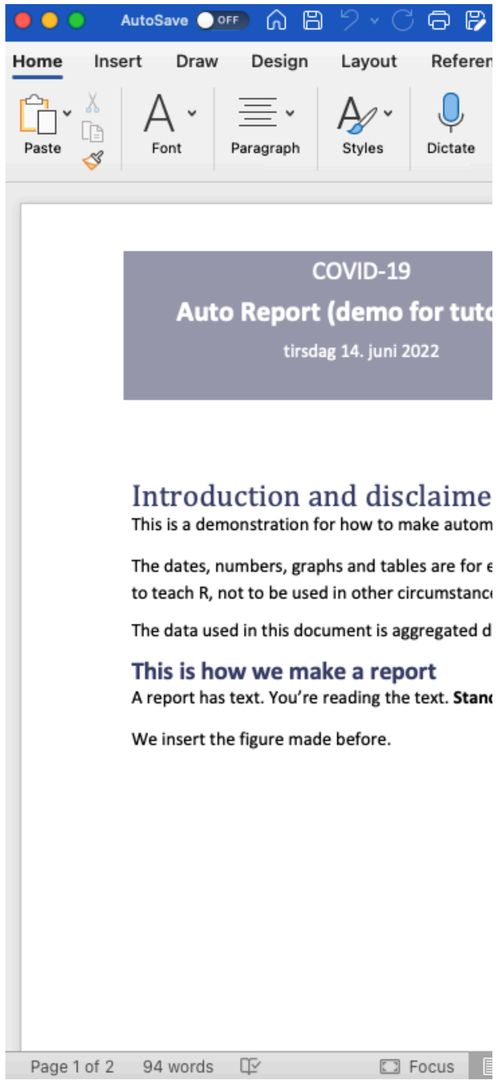
The data used in this document is aggregated data, and it is public available.

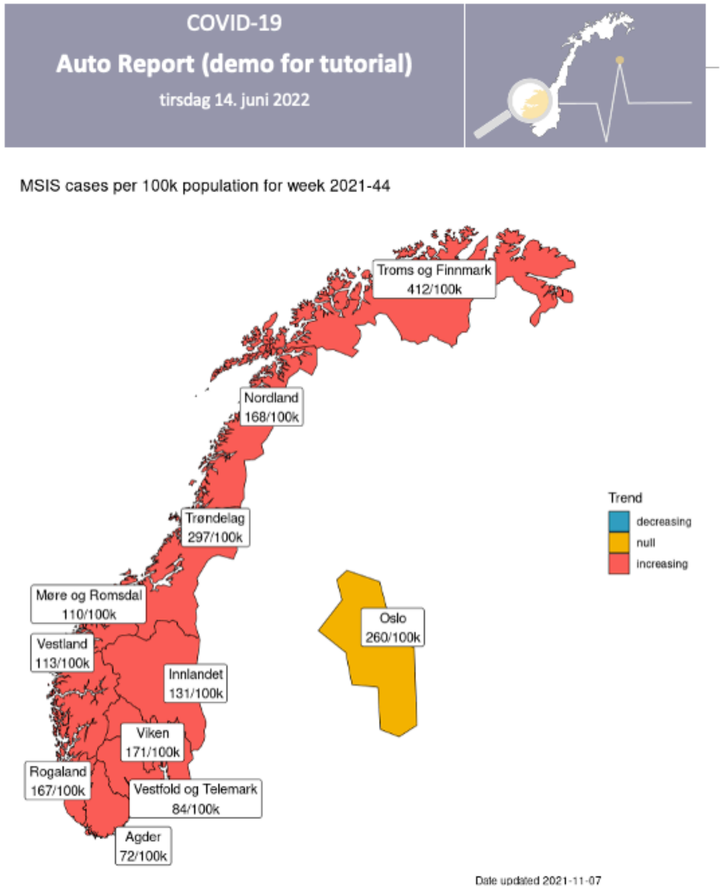
This is how we make a report

A report has text. You're reading the text. **Standard markdown style works here as well**.

We insert the figure made before.

![MSIS Covid-19 cases trend for week 2021-44](map_covid_demo.png)
```





Automated reports

report_demo.Rmd | run_autoreport.R | report_demo_auto.Rmd

Source on Save

```
# set a date ----
today <- as.character(lubridate::today())

# set report path ----
file_path <- './tutorial_autoreport/autoreport_cov

file_output_name <- glue::glue('report_demo_auto_'
file_output_name

file_rmd <- paste0(file_path, 'report_demo_auto.Rm
file_docx <- paste0(file_path, file_output_name)
```

scr

report_demo.Rmd | run_autoreport.R | report_demo_auto.Rmd

Knit | Run

report_auto.Rmd

```
---
title: "A more automated report made on `r today`"
output:
  word_document:
    reference_docx: style_demo.docx
editor_options:
  chunk_output_type: console
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

Introduction and disclaimers

This is a demonstration for how to make automated reports with Rmarkdown and MS Word.

The dates, numbers, graphs and tables are for educational purposes only. The text is made to teach R, not to be used in other circumstances.

The data used in this document is aggregated data, and it is public available.

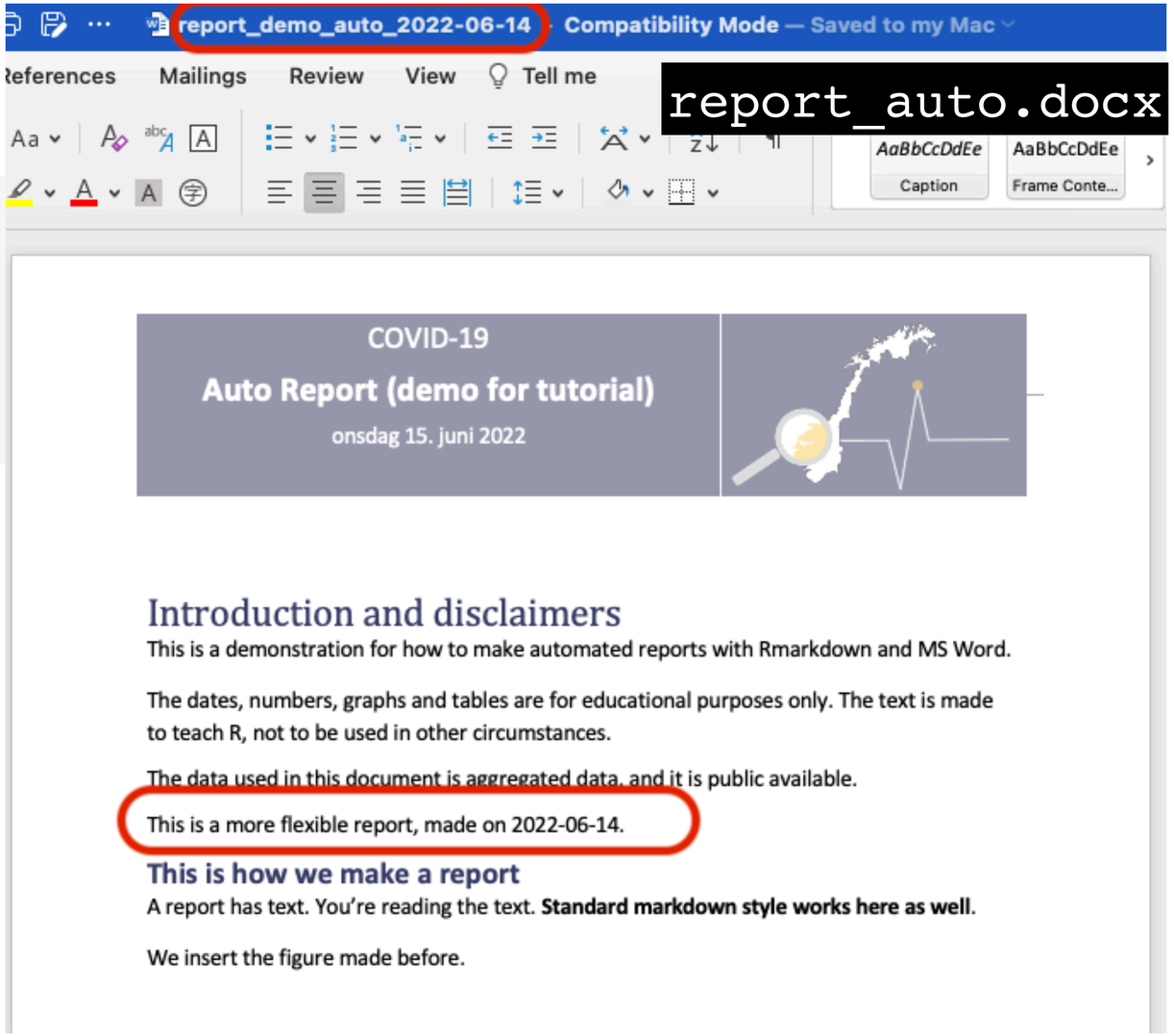
This is a more flexible report, made on `r today`.
```

Automated reports

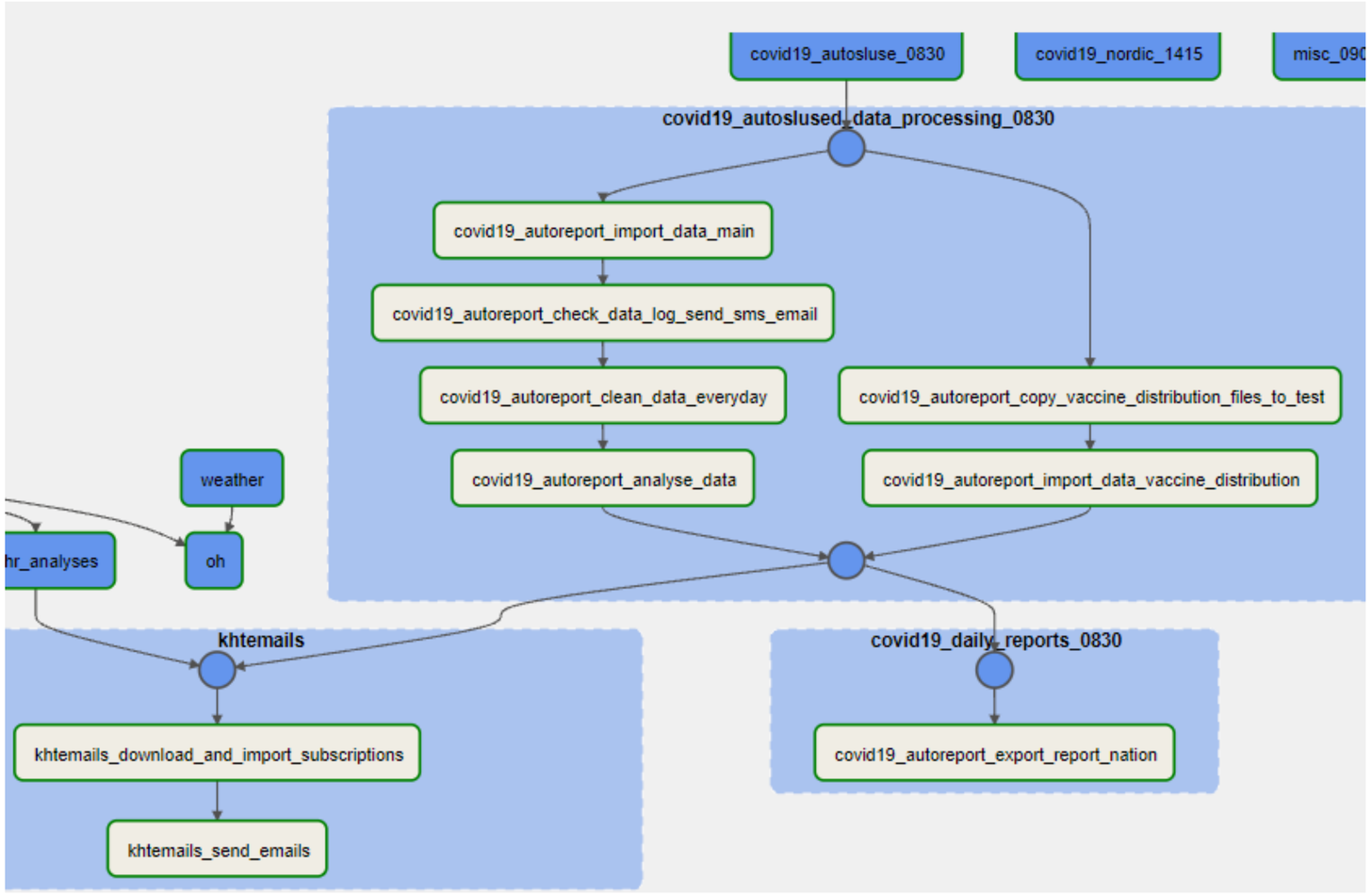
script.R

```
14 # render report ----
15 rmarkdown::render(
16   input = file_rmd,
17   output_dir = file_path,
18   output_file = file_output_name
19 )
```

report_auto.docx



Development, Automation, QC



We run 150+ tasks in airflow

Alert system: Email + SMS

Quality check: vakt
(debug, check report output)

Ways forward

Ways forward

Prevention is better than treatment

Monitoring and surveillance

Infectious diseases in human
Zoonotic diseases (One Health)

Invest in **infrastructure and manufacturing**

Vaccine production and distribution
high protective gears, ⋯

Lab security; training and education; ⋯

Rapid development of tests and vaccine

Collaboration is key!

Data and information sharing, open source software can make it easier !

Consortium for Statistics in Disease Surveillance

2023 ?

Resources

Public health surveillance and preparedness

Centers for Disease Control and Prevention (CDC) guide on public health surveillance https://www.cdc.gov/training/publichealth101/surveillance.html

Book on covid and pandemic (“Preventable” by Devi Sridhar)

Coursera course (JHU) on surveillance https://www.coursera.org/learn/epidemiology-surveillance-systems-analysis/

Our world in data https://ourworldindata.org/coronavirus/country/norway

Johns Hopkins COVID data repository https://github.com/CSSEGISandData/COVID-19

FHI data and reports

Weekly report https://www.fhi.no/publ/2020/koronavirus-ukerapporter/

Statistics bank https://statistikk.fhi.no

CSIDS

Repository of CSIDS https://github.com/csids

Reporting automation https://www.rstudio.com/resources/webinars/rethink-reporting-with-automation/