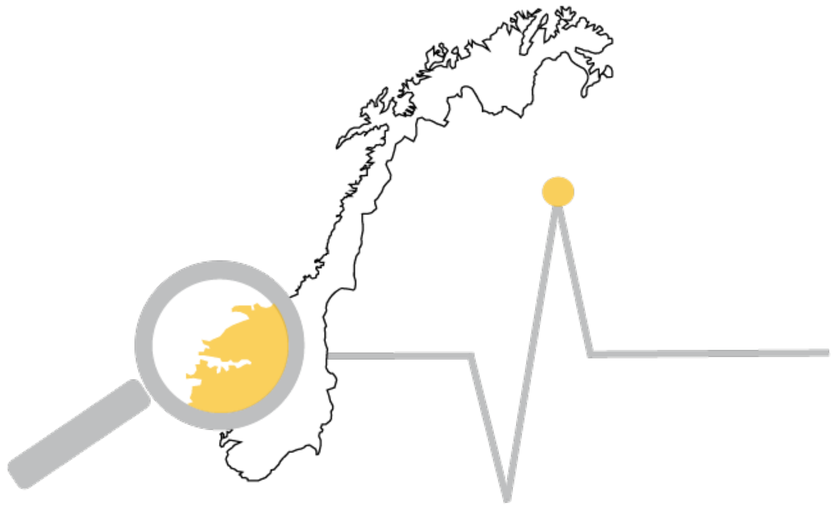
FHI
Folkehelseinstituttet



Introducing Sykdomspulsen

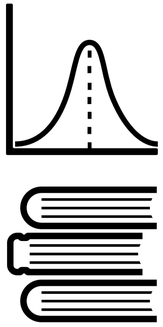
An automated public health surveillance platform

Chi Zhang
@Andreasheen chi.zhang@fhi.no
Norwegian Institute of Public Health (FHI), OCBE

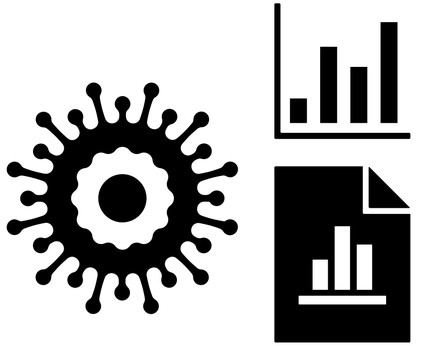
2022-06-16 @useR_Oslo
Github: sykdomspulsen-org
Website: https://docs.sykdomspulsen.no

About me

OCBE–UiO
PhD student



FHI Sykdomspulsen
Researcher



Statistician
+
R developer

2016 2017 2018 2019 2020 2021 2022

Joined Oslo R

1st R talk (blogdown)

2nd R talk (R ladies)

This R talk

About this talk

Public health surveillance
(10 min)

What is PH surveillance
How is the data generated
Surveillance vs Research

Surveillance x Data science = Sykdomspulsen
(10 min)

SP makes reports, and so much more
SP uses open-source tools
SP follows Tidyverse principles

A day at Sykdomspulsen
(20 min)

Example: Covid19 cases report

Splmaps, splstyle
Autoreport tutorial

Sykdomspulsen in production
(5 min)

Task scheduling, automation, CI/CD
Trouble shooting, teamwork

Disclaimers

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

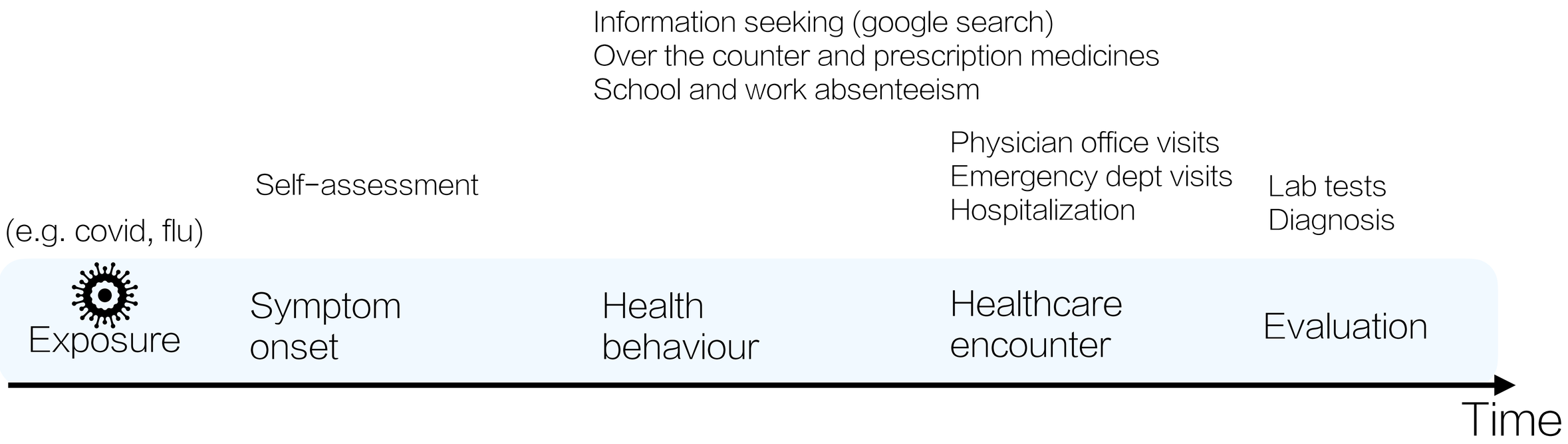
Reports shown in this talk are for educational purposes only.

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

Public health surveillance in a nutshell 

Public health surveillance

Where does data come from?



Data collected by **healthcare professionals**, then some are sent to **regional / national PH authority** for actions

Inspiration: Coursera JHU course on surveillance

Public health surveillance

The 4 Ws and 1H

What

Infectious/communicable diseases
(flu, covid ⋯),

Mortality, causes of death,

Chronic diseases
(cancer, diabetes ⋯),

Meteorological and natural disasters
(heatwave, hurricane),

Others
(firearm injury in US ⋯)

When & Where

Routine (daily, yearly);
Occasional (event related)

Different levels of geo-locations;
Event related

Who

Public Health authorities
(FHI, CDC, PHE ⋯)

How

CADA

Surveillance cycle

Collection,
Analysis,
Dissemination,
Action

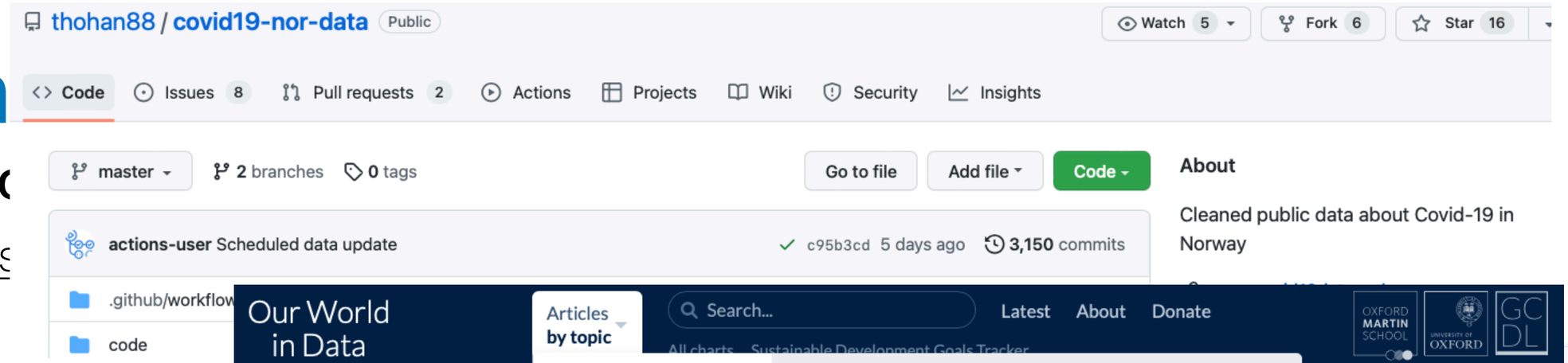
<u>Contribute jointly</u>

Intl organisations (WHO)
NGO (Red Cross)
Academia
Public

Public h

Open data, c

PH authorities



Local Media

VG
Aftenposten
...

Intl Media

Financial Time
Economist
...

Open source projects (GitHub

JHU Covid data repository;
Norway, covid19data repo
Our World In Data
...

Un

in Norway at municipality and district-level reported to the Norwegian Surveillance System for Communicable Diseases (MSIS) with fully available source code.

Norway Case Dashboard

Public health surveillance

Compared to “research”

Surveillance cycle: CADA

Tasks	Research	Surveillance (especially disease surveillance with outbreak potential)	Challenges in surveillance
Collection	Fixed datasets	New data, multiple sources, every day	Data aggregation and cleaning
Analysis	Various	Descriptive (e.g. count, trend)	Flexibility age, sex, time, location groups
Dissemination	(Journal) publication	Reports, graphs, numbers, datasets	Scale and consistency
Action		Policy responses (effect immediately)	Rapid new task development

Surveillance methods can NOT live without research!

Inspiration: Coursera JHU course on surveillance

Surveillance x Data Science tools

= Sykdomspulsen

Sykdomspulsen

The Disease Pulse

A team of 8* members at **Norwegian Institute of Public Health**
Started in 2012. Key role in Covid-19 pandemic

Automated real time public health surveillance platform

Daily** data input

Daily analysis

Daily delivery

Infectious diseases

Covid19

Influenza

Other respiratory

Gastrointestinal

Other infections

Mortality

Excess mortality

Cause of death

Data cleaning and censoring

Statistical analysis

Graph making and reporting

* Regular team members only
** Mortality surveillance is either weekly or yearly

Sykdomspulsen

Automated situational reports

1 nation + 11 counties + 356 municipalities

Everyday, before **7am**

Used by FHI leadership and Ministry of Health

- 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

COVID-19

Dagsrapport – nasjonalt og per fylke

onsdag 20. oktober 2021

Demo

Dagens foreløpige (u.off) tall. Rapporten er generert kl. 1 . 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.

Sykdomspulsen

Automated situational reports

35 pages, 17 tables, 21 figures

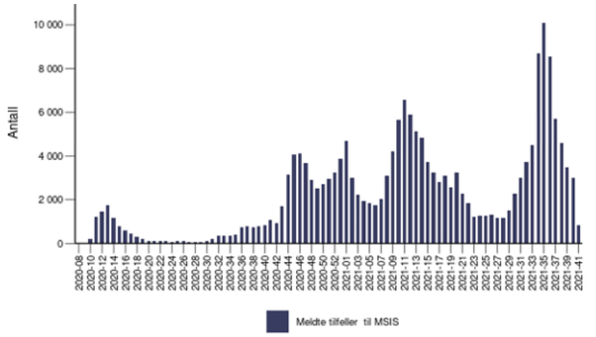
Tabell 1 Covid-19 status og utvikling, uke 2021-37 til uke 2021-41.

Demo

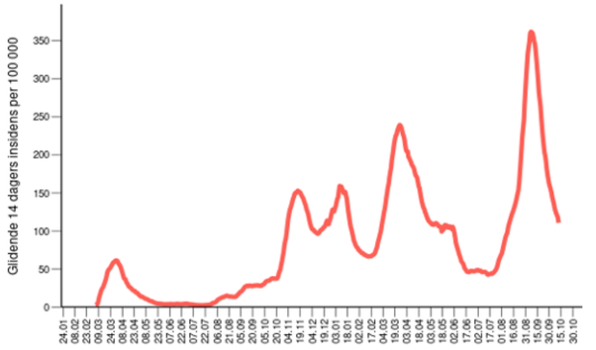
Overvåkingssystem/ Indikatorer	2021-37	2021-38	2021-39	2021-40	2021-41	t av de siste 5 ukene
Meldte tilfeller til MSIS (prøvedato)						
Meldte tilfeller til MSIS (registrertdato)						
Antall personer testet for SARS-CoV-2 (PCR)						
Nye covid-19 positive pasienter innlagt i sykehus (alle årsaker)						
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).



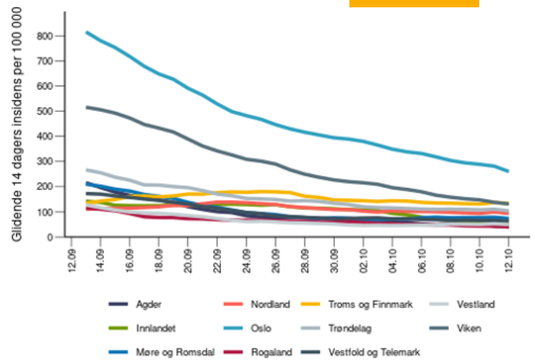
Figur 1. Antall tilfeller av covid-19 per uke basert på prøvedato gjennom hele pandemiperioden, Norge.



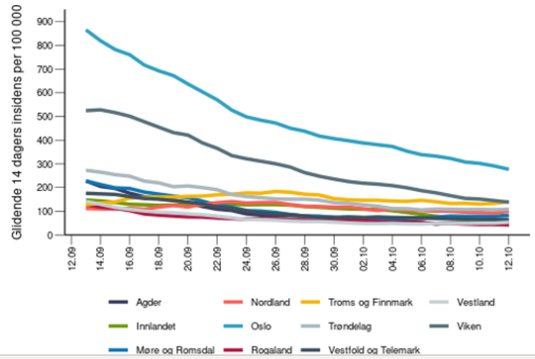
Figur 2. Glidende 14-dagers insidens per 100 000 innbyggere per dag basert på prøvedato gjennom hele pandemiperioden, Norge.

Fylker

Demo



Figur 5. Antall tilfeller per 100 000 innbygger per 14 dager etter fylke og prøvedato, uke 2021-37 til uke 2021-41.



Sykdomspulsen

Automation saves time and cost

Historically (pre 2020.12) made by hand, in the early morning, for selected locations
Overtime estimated by one employee work from 6 to 8am for one year (2h per day)
One employee is allowed 200h overtime per year

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

One year deliverables	Manual report	Sykdomspulsen
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)

Sykdomspulsen

SP produces more than just reports

2 Examples	sKUHR outbreak detection from GP consultations	Excess mortality monitoring
Frequency of updates	Daily	Weekly
Data source	88+ million rows 674 columns	640k+ rows
Analyses	Age group * ICPC2 code * nation/county/municip * years = 100.000+ analyses	Age group * county * years = 1.000+ analyses
Delivery	Municipal PH authorities, via **Shiny website, email reports**	Internal weekly **reports**, **Shiny website**, **Data and reports** to EuroMOMO network

Sykdomspulsen

Deliverables & stakeholders

FHI

Top leadership

Communication

Lab department

Outbreak group

Modelling team

Data registries

MSIS,
sKUHR,
SYSVAK,
NIPaR,
...

We deliver

Actionable intelligence based on statistical analyses of real-time data

via

Situation reports, website, graphs and tables, emails

Government

National government, Ministry of Health

County government (statsforvalter)

Local PH authority (kommunelege, fylkeslege)

Public

Github
fhi.no
Media (VG, Aftenposten)

International health networks (EuroMOMO, ECDC, WHO)

Sykdomspulsen

The SP way of data science

Tools, Team, Rules

Sykdomspulsen

The tools

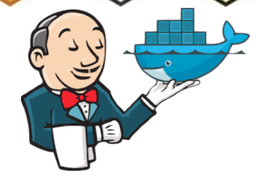
Designed and developed by **statisticians**

Free, open-source solutions

Flexible, fast development
(13 R packages developed in-house)

RStudio

GitHub



Apache Airflow

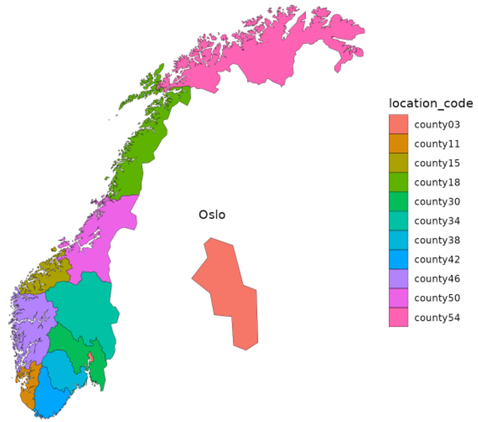
```
fhidata::norway_locations_names()
#>      location_code   location_name
#>   1:         norge           Norge
#>   2:      county42           Agder
#>   3:      county34       Innlandet
#>   4:      county15 Møre og Romsdal
#>   5:      county18        Nordland
#>  ---
#> 581:     faregion4            Midt
#> 582:     faregion5            Nord
#> 583:     faregion2       Stor-Oslo
#> 584:     faregion3     Sør og vest
#> 585:     faregion1             Øst
```

Customise the maps

The maps can be easily customized with color and labels.

Automatic coloring by `location_code`

```
pd <- copy(fhimaps::norway_nuts3_map_b2020_insert_oslo_dt)

q <- ggplot()
q <- q + geom_polygon(
  data = pd,
  mapping = aes(
    x = long,
    y = lat,
    group = group,
    fill = location_code
  ),
  color="black",
  size=0.1
)
q <- q + annotate(
  "text",
  x = fhimaps::norway_xxx_position_title_insert_oslo_b2020_insert_oslo_dt$lon
  y = fhimaps::norway_xxx_position_title_insert_oslo_b2020_insert_oslo_dt$lat
  label = "Oslo"
)
q <- q + theme_void()
q <- q + coord_quickmap()
q <- q + labs(title = "")
q
```



Sykdomspulsen

The team

Project manager	Coordinator	Web designer	Statistician / developer	Engineer
x1	x1	x1	x4	x1

SP team uses R and Git

Monthly training (“Tuesday learning hour”)

Debug training and Vakt week

“Engineer of the week”

(Regular team members only)

Sykdomspulsen

The rules

SP adopts **tidyverse principles**

Tidyverse principles	SP practice
Human centered	We use R. Code for human, easy to understand
Consistent	Naming system, data format, e.g. `msis_by_time_location_age_sex_2022-06-15`
Composable	Code can be broken down into pieces (‘chunking’), easy to scale up and debug
Inclusive	Open-source tools, training, sharing

Doing everyone’s data work

->

Build tools that enable everyone to do their data work faster

Sykdomspulsen

Tools revisited: splverse x surveillance

Tasks	Challenges	Solution
Collection	Data aggregation and cleaning	
Analysis	Flexibility age, sex, time, location groups	
Dissemination	Scale and consistency	
Action	Rapid new task development	

A day at Sykdomspulsen

A day at Sykdomspulsen

We received a request

How is the situation of **Covid trend** in Norway and 11 counties?
We need a report with **map**, **everyday**.

Ask questions:

- What **data** should be used?
- What **specifications**? Age and sex, location, time
- What **metrics and analysis**? Number of cases (incidence/ cumulative)? Percentage? Trend? Excess? Outbreak?
- What kind of **delivery**? Report, website, tables, graphs, emails …
- When is the deadline? **How often**?

Data: Covid cases (MSIS registry)
Specification: All age, both sex, national + county
Metrics and analysis: # new cases, /100k, weekly trend
Delivery: Graph, map
Delivery frequency: Daily

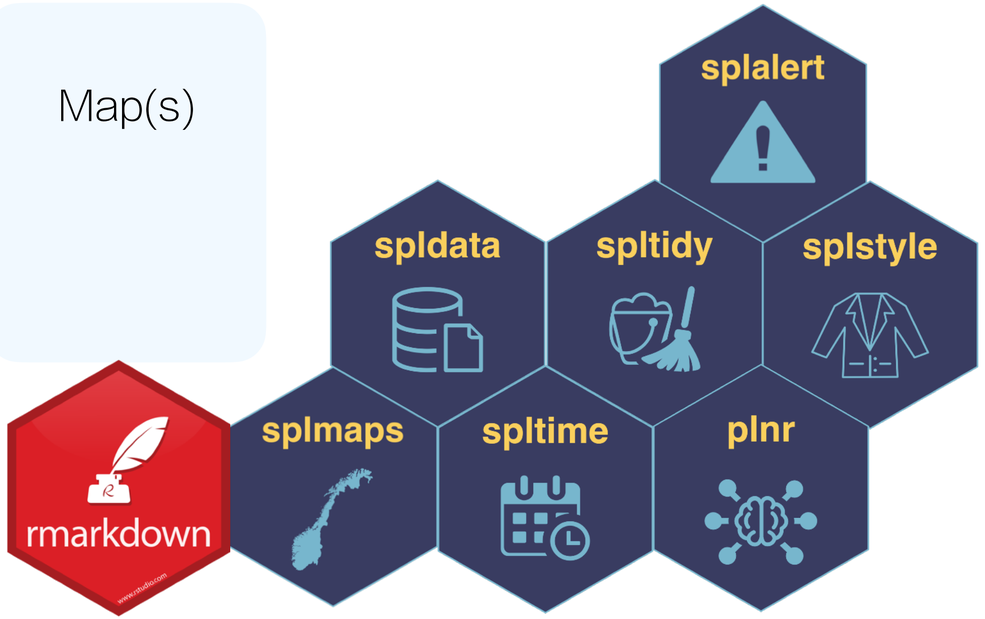
MSIS: Norwegian Surveillance Systems for Communicable Diseases

Daily report for Covid trend (MSIS)

Task breakdown: analysis + reporting

Load data Clean Standardize	EDA Initial visualize	Analysis for ONE location (or, age, sex, outcome ⋯)	Extend to Multiple locations	Map(s)

Make one report -> Automate multiple reports



Daily report for Covid trend (MSIS)

Step 1: data

Demo data is pre-formatted, publicly available
Censoring rules apply

```
d_msis <- spltidy::covid19_msis_cases_b
# check the column names
colnames(d_msis)
```

```
 [1] "granularity_time"
 [2] "granularity_geo"
 [3] "country_iso3"
 [4] "location_code"
 [5] "border"
 [6] "age"
 [7] "sex"
 [8] "isoyear"
 [9] "isoweek"
[10] "isoyearweek"
[11] "season"
[12] "seasonweek"
[13] "calyear"
[14] "calmonth"
[15] "calyearmonth"
[16] "date"
[17] "covid19_cases_testdate_n"
[18] "covid19_cases_testdate_pr100000"
```

fixed fields, consistency

task specific

	[unified]	[unified]	[unified]	[unified]	[unified]
	<character>	<character>	<numeric>	<integer>	<integer>
	NA=0 %	NA=0 %	NA=0 %	NA=13 %	NA=13 %
	isoyearweek	**season**	**seasonweek**	**calyear**	**calmonth**
1:	2020-08	2019/2020	31	2020	2
2:	2020-08	2019/2020	31	2020	2
3:	2020-08	2019/2020	31	2020	2
4:	2020-09	2019/2020	32	2020	2
5:	2020-09	2019/2020	32	2020	2
...					
11024:	2022-14	2021/2022	37	NA	NA
11025:	2022-15	2021/2022	38	NA	NA
11026:	2022-16	2021/2022	39	NA	NA
11027:	2022-17	2021/2022	40	NA	NA
11028:	2022-18	2021/2022	41	NA	NA

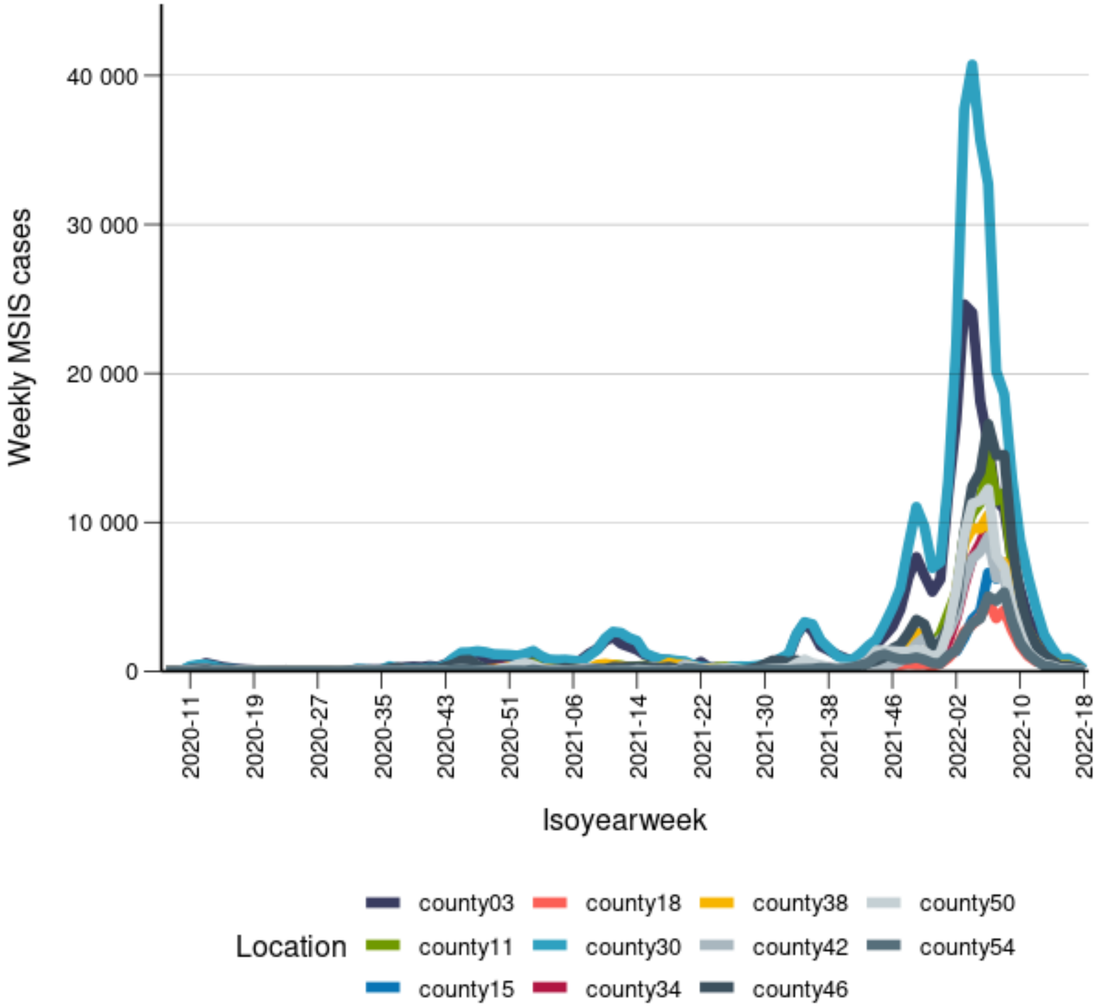
	[unified]	[unified]	[context]	[context]
	<character>	<Date>	<integer>	<numeric>
	NA=13 %	NA=0 %	NA=0 %	NA=0 %
	calyearmonth	**date**	**covid19_cases_testdate_n**	**covid19_cases_testdate_pr100000**
1:	2020-M02	2020-02-21	0	0
2:	2020-M02	2020-02-22	0	0
3:	2020-M02	2020-02-23	0	0
4:	2020-M02	2020-02-24	0	0
5:	2020-M02	2020-02-25	0	0
...				
11024:	NA	2022-04-10	6888	126.9614
11025:	NA	2022-04-17	3635	67.0013
11026:	NA	2022-04-24	3764	69.379
11027:	NA	2022-05-01	2243	41.3436
11028:	NA	2022-05-08	502	9.253

Daily report for Covid trend (MSIS)

Step 2: select indicator, exploratory visualise

Visualize weekly cases for 11 counties

```
d_fylke <- copy(d_msis[granularity_time == 'isoweek'& 
'county'] )


q <- ggplot(d_fylke, aes(x = isoyearweek, y = covid19_
q <- q + geom_line(aes(group = location_code, col = lo
= 2)
q <- q + scale_y_continuous("Weekly MSIS cases",
                            expand = expansion(mult = 
                            labels = splstyle::format 
q <- q + scale_x_discrete("Isoyearweek",
                          breaks = splstyle::every_nth

q <- q + splstyle::scale_color_fhi("Location", palette
q <- q + splstyle::theme_fhi_lines_horizontal(legend_p
"bottom")
q <- q + splstyle::set_x_axis_vertical()
q
```

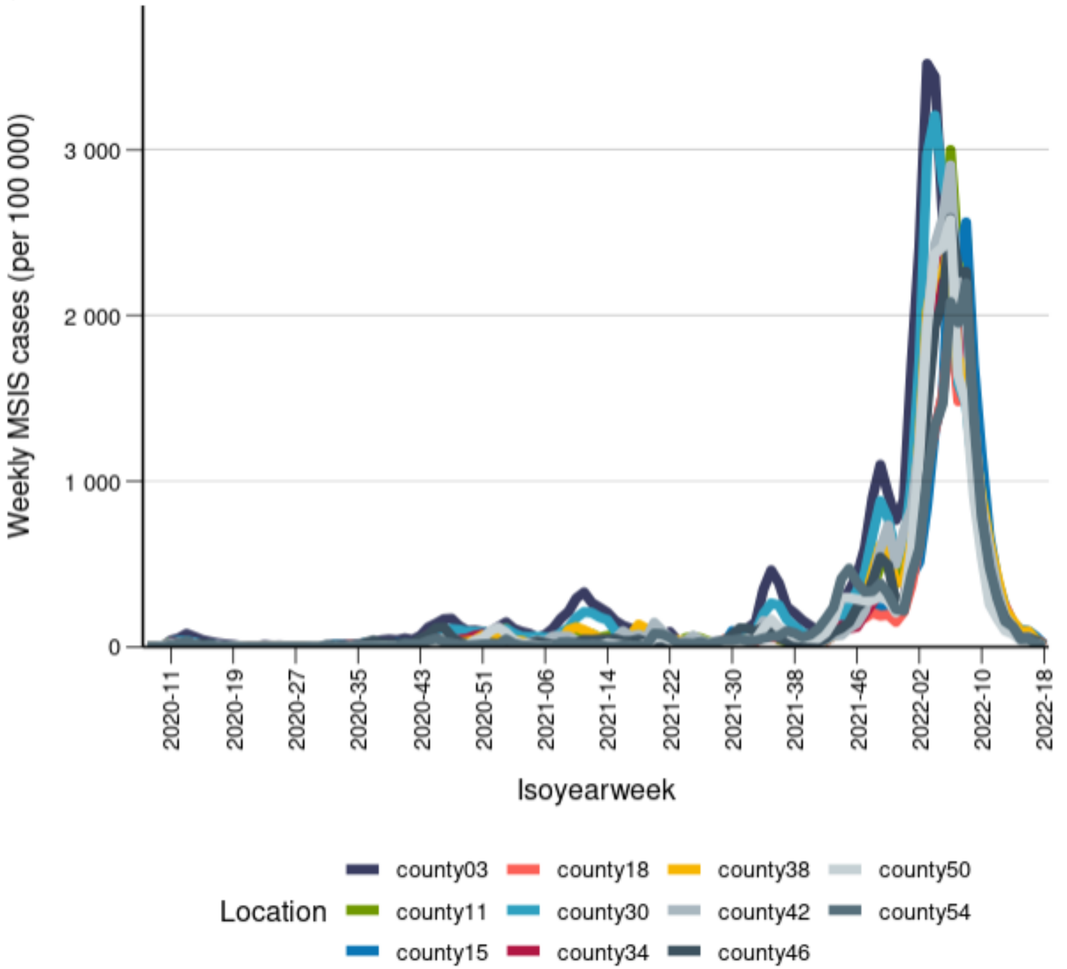


Daily report for Covid trend (MSIS)

Step 2: select indicator, exploratory visualise

Visualize weekly cases per 100k population for 11 counties

```
q <- ggplot(d_fylke, aes(x = isoyearweek, y =
covid19_cases_testdate_pr100000))
q <- q + geom_line(aes(group = location_code, col = l
= 2)
q <- q + scale_y_continuous("Weekly MSIS cases (per 1
                            expand = expansion(mult =
                            labels = splstyle::format

q <- q + scale_x_discrete("Isoyearweek",
                          breaks = splstyle::every_nt

q <- q + splstyle::scale_color_fhi("Location", palett
q <- q + splstyle::theme_fhi_lines_horizontal(legend_
"bottom")
q <- q + splstyle::set_x_axis_vertical()
q
```



Daily report for Covid trend (MSIS)

Step 3: analysis for one location, visualise with information

Trend analysis (6 weeks) at **national** level
Details on implementation see `splalert` package

```
d_msis_norge <- copy(d_msis[granularity_time == 'isoweek' &
                        isoyearweek >= '2021-26' &
                        location_code == 'norge'])

trend_msis_norge <- splalert::short_term_trend(
  d_msis_norge,
  numerator = "covid19_cases_testdate_n",
  trend_isoweeks = 6,
  remove_last_isoweeks = 1
)
```

splalert 2022.6.15 Get started Reference Articles Changelog

Determine the short term trend

Source: R/short_term_trend.R

Determine the short term trend

Usage

```
short_term_trend(
  x,
  numerator,
  denominator = NULL,
  prX = 100,
  trend_days = 42,
  remove_last_days = 0,
  forecast_days = trend_days,
  trend_isoweeks = ceiling(trend_days/7),
```

sykdomspulsen-org / splalert Public

generated from sykdomspulsen-org/template-r-package

Edit Pins Watch

Code Issues Pull requests Actions Wiki Security Insights

915b98dee1 splalert / R / short_term_trend.R

andreaczhang test short term trend, fix some check issues Latest

2 contributors

404 lines (340 sloc) 15.1 KB

```
1  gen_data_short_term_trend <- function(){
2
```

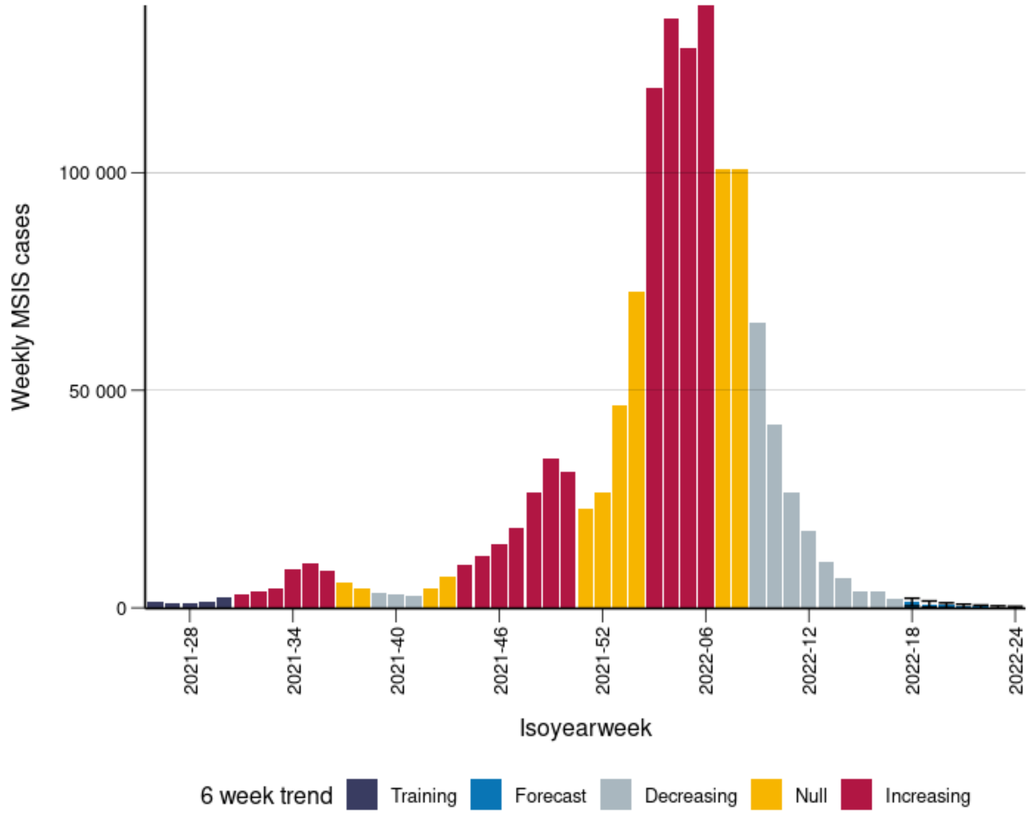
…r/denominator?

Daily report for Covid trend (MSIS)

Step 3: analysis for one location, visualise with information

Trend analysis (6 weeks) at national level

```
# plot
q <- ggplot(trend_msis_norge, aes(x = isoyearweek,
                                  y = covid19_cases_testdate_forecasted_n,
                                  group = 1))
q <- q + geom_col(mapping = aes(fill =
covid19_cases_testdate_trend0_42_status ))
q <- q + geom_errorbar(
  mapping = aes(
    ymin = covid19_cases_testdate_forecasted_predinterval_q02x5_n,
    ymax = covid19_cases_testdate_forecasted_predinterval_q97x5_n
  )
)
q <- q + scale_y_continuous("Weekly MSIS cases",
                            expand = c(0, 0.1),
                            labels = splstyle::format_nor_num_0)
q <- q + scale_x_discrete("Isoyearweek", breaks =
splstyle::every_nth(6))
q <- q + expand_limits(y=0)
q <- q + splstyle::scale_fill_fhi("6 week trend", palette = "contrast")
q <- q + splstyle::theme_fhi_lines_horizontal(legend_position =
"bottom")
q <- q + splstyle::set_x_axis_vertical()
q
```

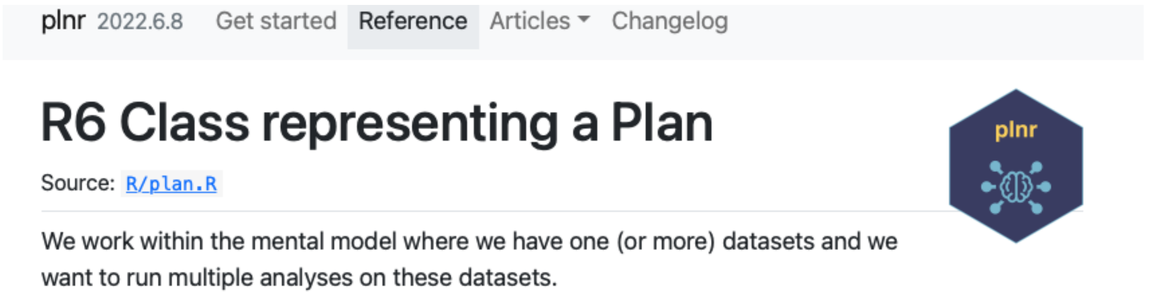


Daily report for Covid trend (MSIS)

Step 4: analyses for all locations

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 code for one subset
- repeat for all combinations



Daily report for Covid trend (MSIS)

Step 4: analyses for all locations

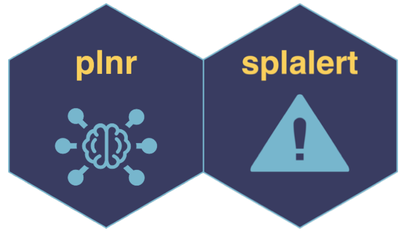
```r
# We begin by defining a new plan
p <- plnr::Plan$new()

# 1. add data
data_fn <- function(){
  return(d_msis)
}

p$add_data(
  name = "covid19 cases",
# 2. add argset
# check location codes
location_codes <- p$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()
```

```
> location_codes <- p$get_data()$covid19_cases$location_code %>%
+   unique() %>%
+   print()
 [1] "county03" "county11" "county15" "county18"
 [5] "county30" "county34" "county38" "county42"
 [9] "county46" "county50" "county54" "norge"
```

```
> p$get_argsets_as_dt()
                          name_analysis index_analysis location_code granularity_time
 1: 8138784e-5338-41c8-9bc7-d02d076b5cd9              1      county03          isoweek
 2: 192d06ba-0b16-4dbf-87a1-0050f4097ae6              2      county11          isoweek
 3: 6e1f407e-4e13-41ca-8405-1c444dab23ee              3      county15          isoweek
 4: 78f2038e-77cd-49c0-8503-a1f093daf5d6              4      county18          isoweek
 5: 5cc688cb-b974-4b3f-9e99-37b83dc5920e              5      county30          isoweek
 6: eaa12ba5-d340-4eaf-a871-3728a8e799f5              6      county34          isoweek
 7: 5246bf8c-e4f4-419d-900d-a2c815ee37bb              7      county38          isoweek
 8: 8c45d3e0-486e-4ca7-a1ac-8ef58ac27d8a              8      county42          isoweek
 9: c48c46c0-a387-4e42-943c-c785ba29c78a              9      county46          isoweek
10: 6cd25991-475b-43d2-b286-143f9d6c8cd5             10      county50          isoweek
11: bf828536-33bc-4b9a-a162-471cd78835aa             11      county54          isoweek
12: dad510a2-5c0e-4e44-b352-e62fa5a798e3             12         norge          isoweek
```

Daily report for Covid trend (MSIS)

Step 4: analyses for all locations




```r
# 3. action
# To do this, we first need to create an action function
# (takes two arguments -- data and argset)

action_fn <- function(data, argset){
  if(plnr::is_run_directly()){
    data <- p$get_data()
    argset <- p$get_argset(1)  # county03, isoweek
  }
  # develop function for ONE argset only

  # data
  pd <- data$covid19_cases[
    location_code == argset$location_code &
      granularity_time == argset$granularity_time
  ]

  # function
  trend_msis <- splalert::short_term_trend(
    pd,
    numerator = "covid19_cases_testdate_n",
    trend_isoweeks = 6,
    remove_last_isoweeks = 1
  )
  # return
  trend_msis

}
```

```r
# apply this to all 12 argsets
p$apply_action_fn_to_all_argsets(fn_name = "action_fn")
```

```r
# run together
res <- p$run_all_progress()
```

```
> res <- p$run_all_progress()
[==================================================] 12/12 (100%) in 00:00:05, eta:  0s
```

Can be paralleled

Daily report for Covid trend (MSIS)

Step 5: make a map


```
res_unlisted <- rbindlist(res)
```

```
# select data for one week (narrow data)
this_isoyearweek <- '2021-44'
d_msis_this_isoyearweek <- res_unlisted[granularity_geo == 'county' &
                                        isoyearweek == this_isoyearweek,
                                        .(location_code,
                                          date,
                                          isoyearweek,
                                          covid19_cases_testdate_pr100000,
```

```
covid19_cases_testdate_trend0_4
```

```
d_msis_this_isoyearweek[, trend
  covid19_cases_testdate_trend0
  levels = c('decreasing','null
)]
```

```
> d_msis_this_isoyearweek
```

	location_code	date	isoyearweek	covid19_cases_testdate_pr100000	covid19_cases_testdate_trend0_42_status
1:	county03	2021-11-07	2021-44	260.25452	null
2:	county11	2021-11-07	2021-44	167.41083	increasing
3:	county15	2021-11-07	2021-44	109.58636	increasing
4:	county18	2021-11-07	2021-44	167.67563	increasing
5:	county30	2021-11-07	2021-44	170.63457	increasing
6:	county34	2021-11-07	2021-44	130.86780	increasing
7:	county38	2021-11-07	2021-44	84.14675	increasing
8:	county42	2021-11-07	2021-44	71.55739	increasing
9:	county46	2021-11-07	2021-44	112.86417	increasing
10:	county50	2021-11-07	2021-44	297.16168	increasing
11:	county54	2021-11-07	2021-44	411.69766	increasing

Daily report for Covid trend (MSIS)

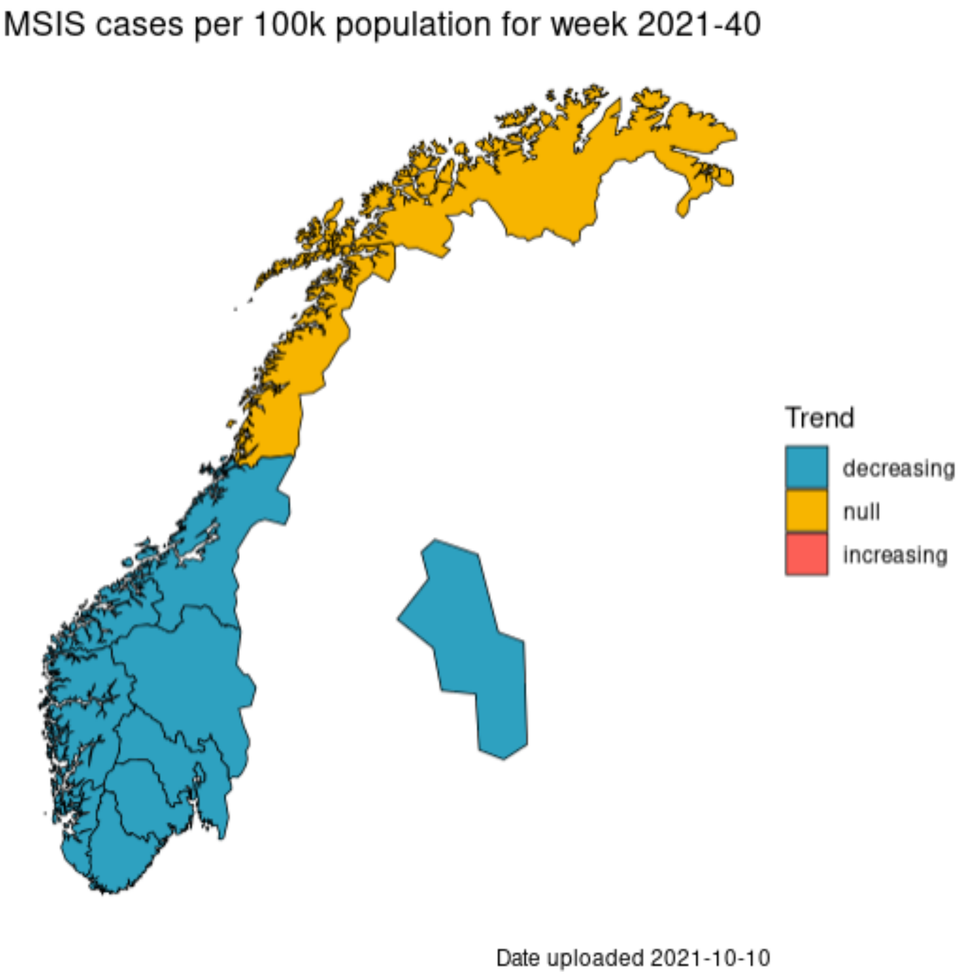
Step 5: make one map

```
pd <- copy(splmaps::nor_nuts3_map_b2020_insert_oslo_dt)


# join the map data.table
pd[d_msis_this_isoyearweek,on="location_code",trend:=trend]

# plot map
q <- ggplot()
q <- q + geom_polygon(
  data = pd,
  mapping = aes(x = long, y = lat, group = group,fill=trend),
  color="black",
  size=0.25
)
q <- q + coord_quickmap()
q <- q + theme_void()
q <- q + labs(title=glue::glue("MSIS cases per 100k population for week
", this_isoyearweek))
q <- q + splstyle::scale_fill_fhi("Trend",palette = "warning", direction
= 1, drop=F)

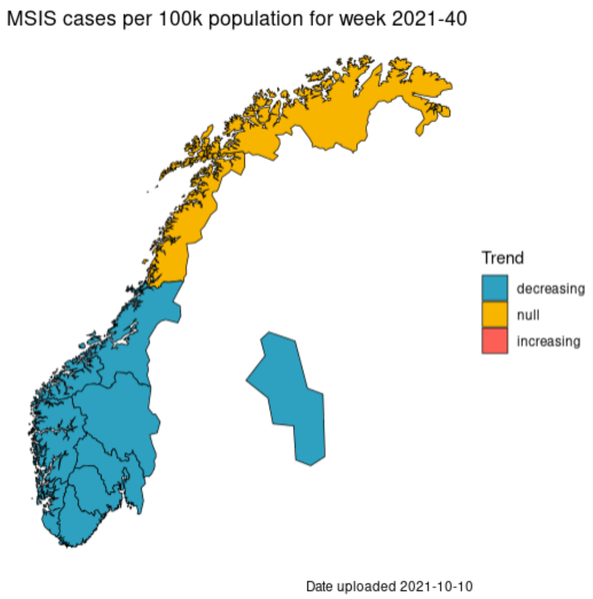
# include date
date_update <- as.character(unique(d_msis_this_isoyearweek$date))
q <- q + labs(caption = glue::glue("Date updated ", date_update))
q
```

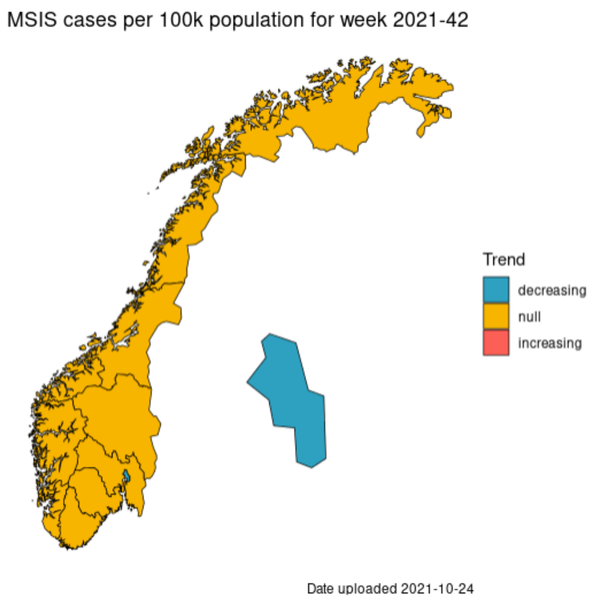


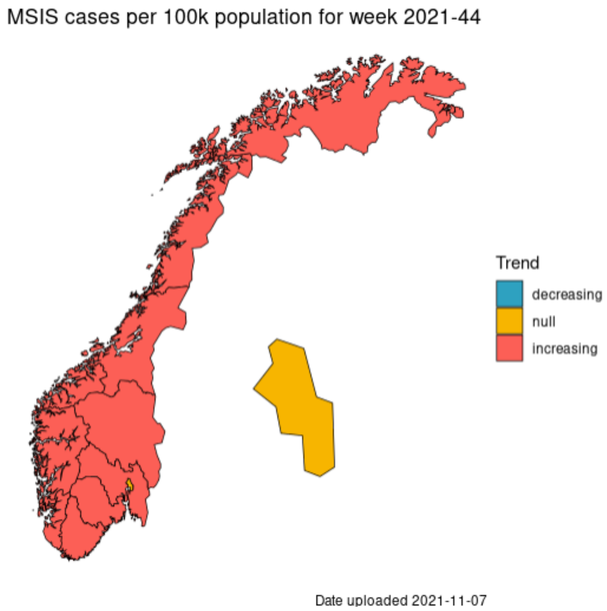
Daily report for Covid trend (MSIS)

Step 5: make some more maps

Trend dynamics for week 40, 42, 44 in 2021 (MSIS confirmed cases)







Daily report for Covid trend (MSIS)

Step 5: make some more maps

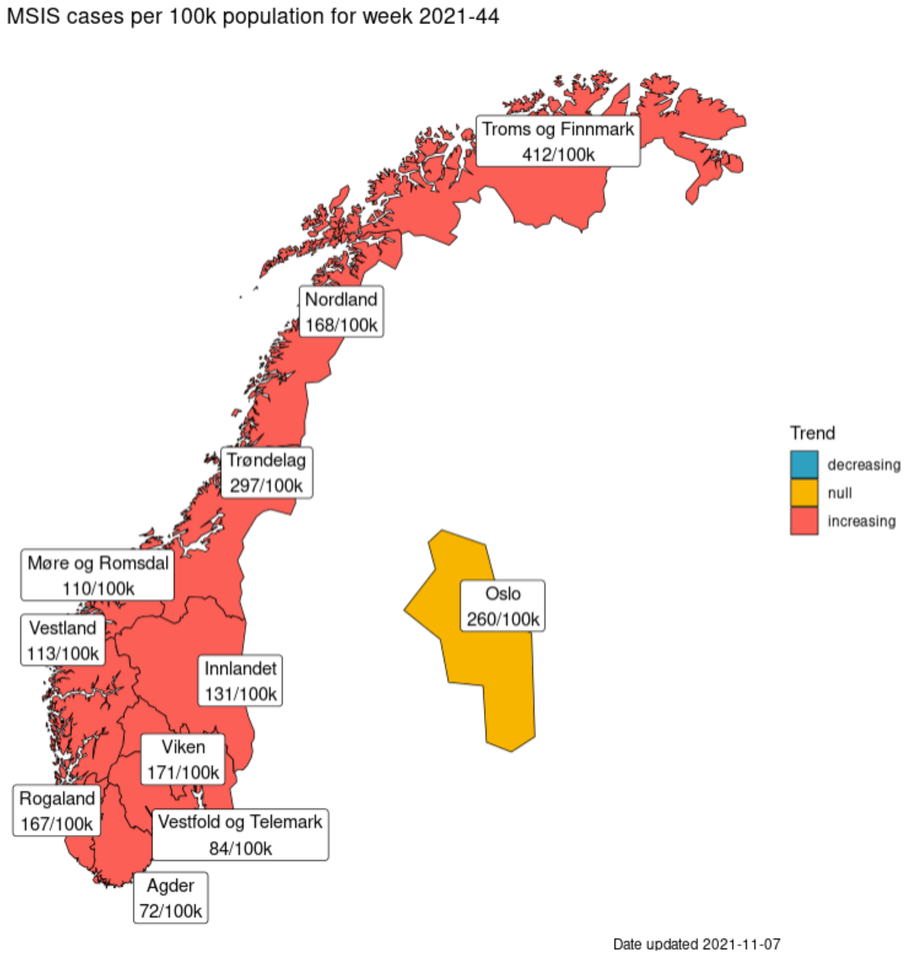
```
# now we add label to the map
# which contains county name + case per 100k
labels <- copy(splmaps::nor_nuts3_position_geolabels_b2020_insert_oslo_d
t)
labels[
  d_msis_this_isoyearweek,
  on = "location_code",
  cases_100k := covid19_cases_testdate_pr100000
]

labels[
  spldata::norway_locations_names_b2020,
  on = "location_code",
  location_name := location_name
]

# format case, remove decimal
labels[, cases_100k := splstyle::format_nor_num_0(cases_100k)]
labels

# put case 100k together with county name
labels[, label := paste0(location_name, '\n', cases_100k, '/100k')]

q <- q + ggrepel::geom_label_repel(
  data = labels,
  mapping = aes(x = long, y = lat, label = label)
)
q
```



Daily report for Covid trend (MSIS)

Step 6: make a (static) report

Requires

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

Input

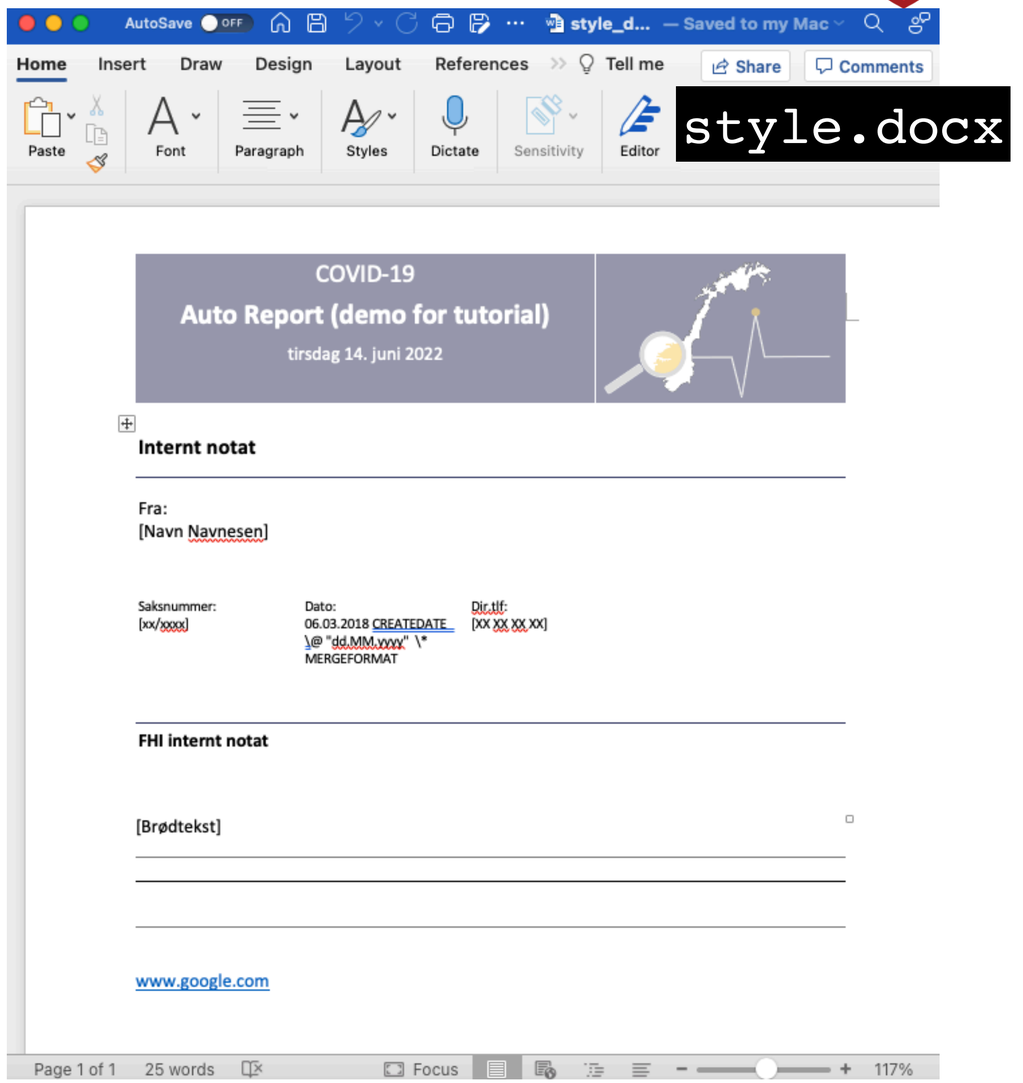
- `report.docx`

Output

- `script.R`

Automate

(Tutorial available!)



Daily report for Covid trend (MSIS)

Step 6: make a (static) report

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.

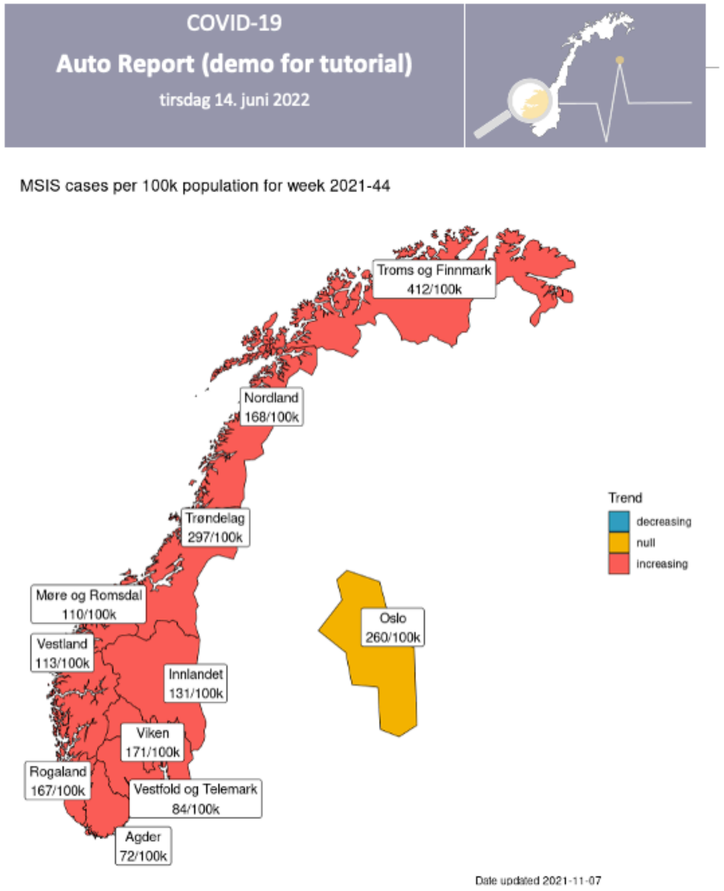
The dates, numbers, graphs and tables are for educational purposes only. The text is made to teach R, not to be used in other circumstances.

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)
```

COVID-19

Auto Report (demo for tutorial)

tirsdag 14. juni 2022

Introduction and disclaime

This is a demonstration for how to make autom

The dates, numbers, graphs and tables are for e

to teach R, not to be used in other circumstanc

The data used in this document is aggregated d

This is how we make a report

A report has text. You're reading the text. Stanc

We insert the figure made before.



Daily report for Covid trend (MSIS)

Step 7: make many autoreports

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)

```

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`.

```

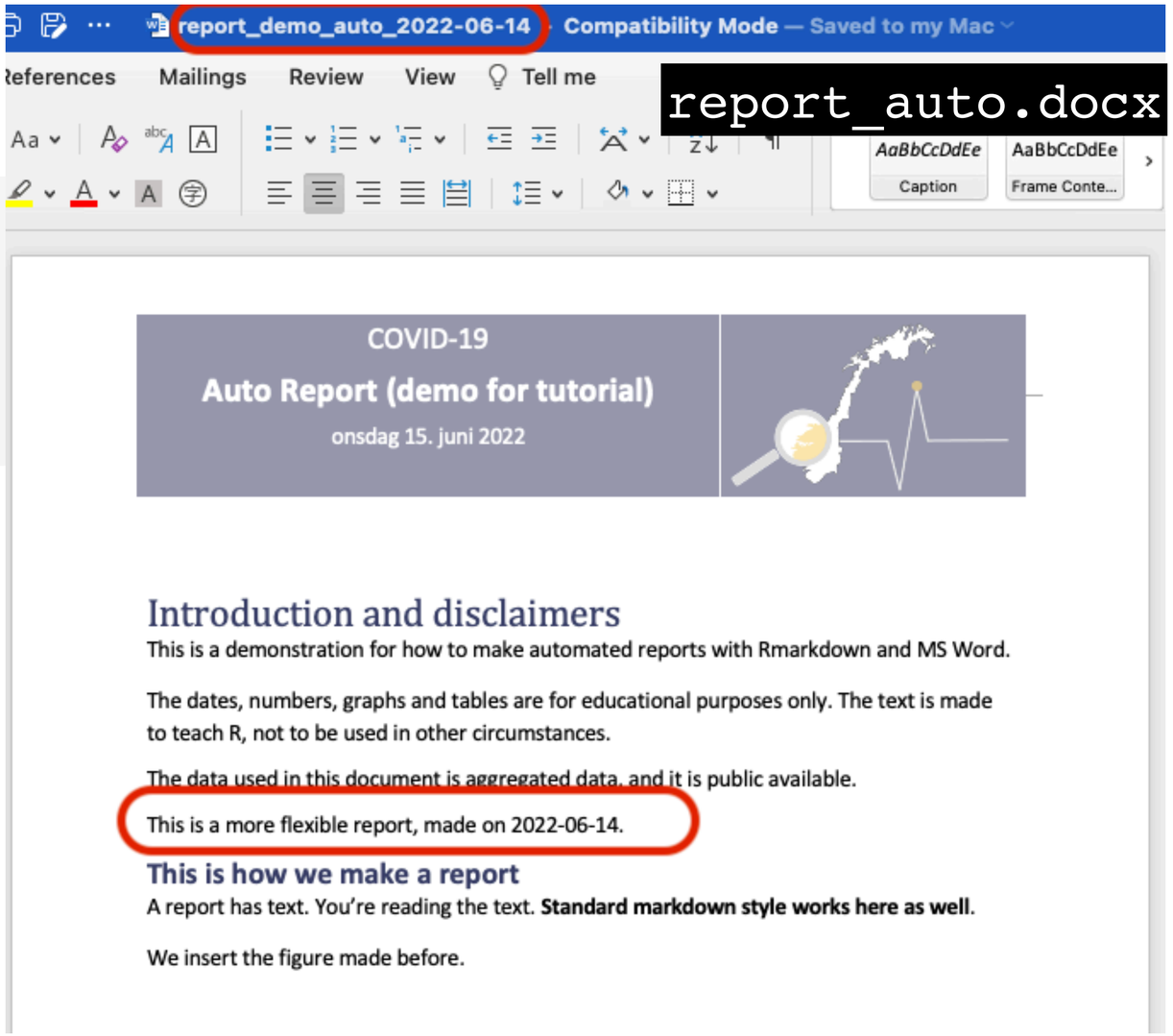
# Daily report for Covid trend (MSIS)

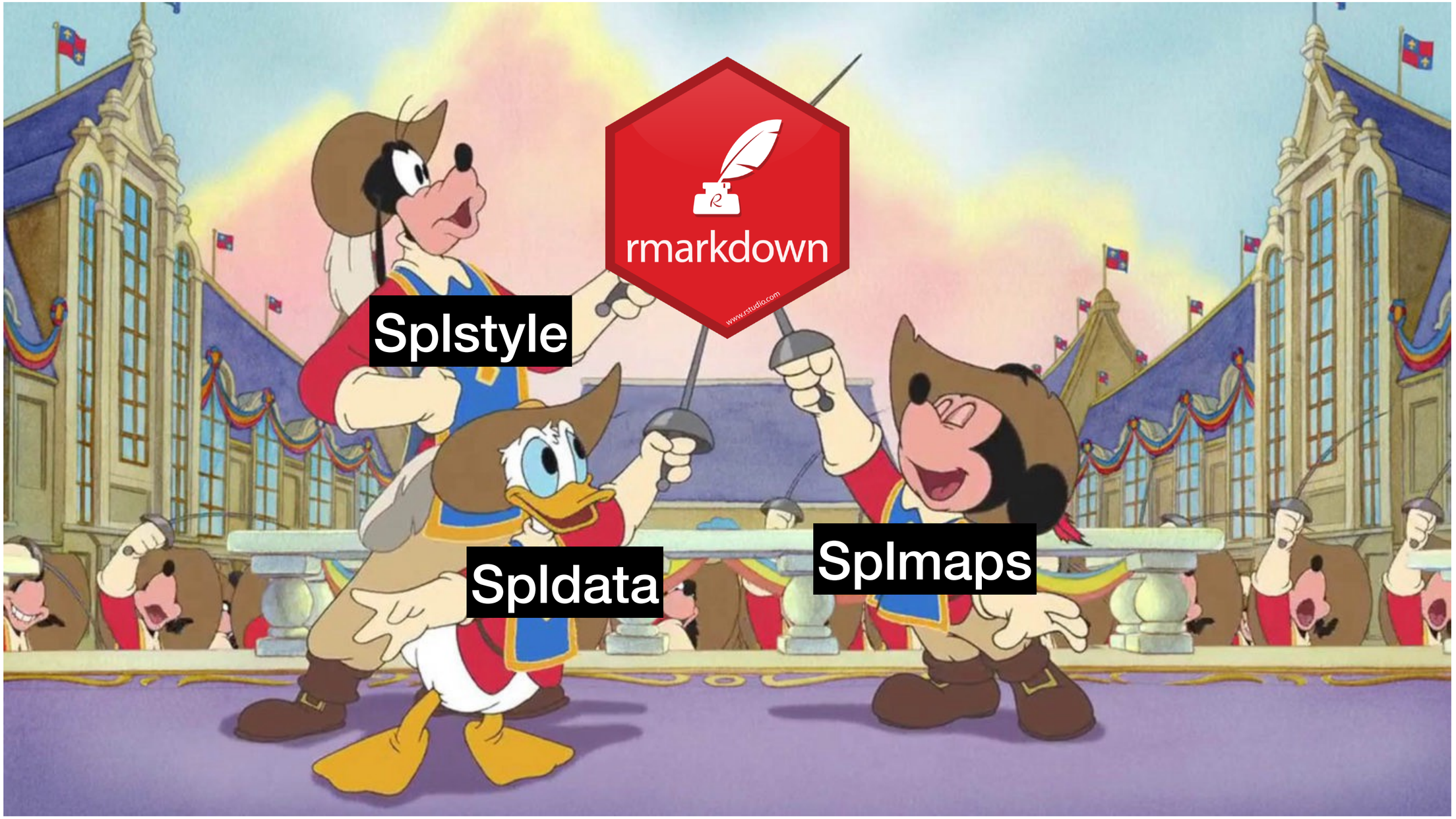
## Step 7: make many autoreports

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




rmarkdown
www.rstudio.com
SpIstyle
SpIdata
SpImaps

Bonus time!

Bonus 1: SP packages

splmaps 2022.6.14 Get started Reference Articles ▾ Changelog

splmaps

Overview

splmaps helps create preformatted ma
helps public health reporting purposes

The following geographic granularities

- County
- Municipality
- City ward (only for Oslo)

It is convenient to visualize maps with
palettes.

Read the introduction vignette here or

Customized coloring with external information

It is also possible to specify the color by user-defined groups. Here we show an example of assigning different (pseudo) risk level to each county.

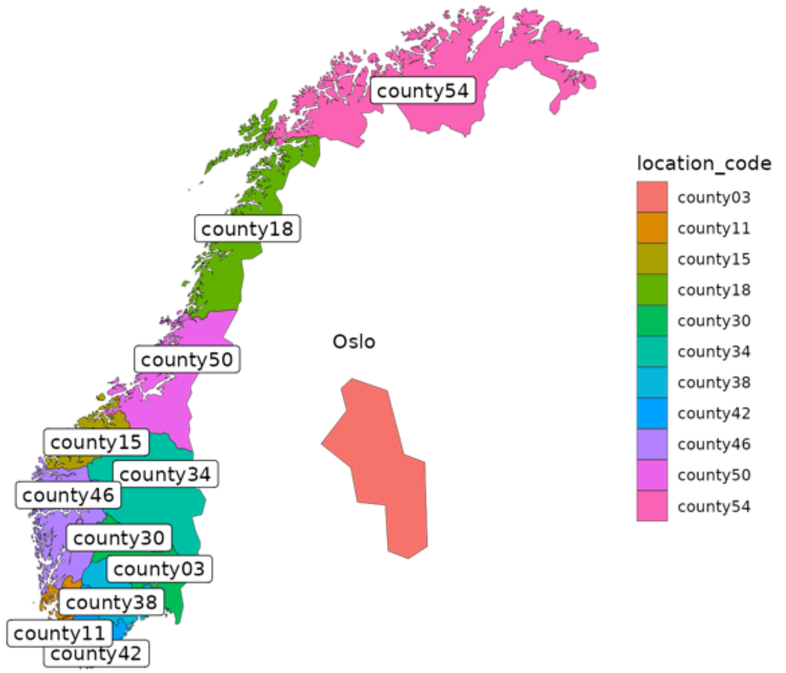
```
pd <- copy(splmaps::nor_nuts3_map_b2020_insert_oslo_dt)

# assign each location a random category for different colors
location_info <- unique(pd[,c("location_code")])
location_info[,category:=rep(
  c("Good","Normal","Neutral","Bad","Very Bad"),
  each=3)[1:.N]
]
location_info[,category:=factor(
  category,
  levels=c("Good","Normal","Neutral","Bad","Very Bad")
  )
]
print(location_info)
#>     location_code category
#>  1:      county11     Good
#>  2:      county15     Good
#>  3:      county18     Good
#>  4:      county03   Normal
#>  5:      county30   Normal
#>  6:      county34   Normal
#>  7:      county38  Neutral
#>  8:      county42  Neutral
#>  9:      county46  Neutral
#> 10:      county50      Bad
#> 11:      county54      Bad
```

On this page

Colored maps

Automatic coloring by location_code



Bonus 1: SP packages

splstyle 2022.5.31 Get started Reference Articles Changelog

- Tables
- Epicurves
- Maps with inset
- Palettes
- Plot with scaled y-axis
- Themes

splstyle

Overview

splstyle contains helpful functions for creating sed by FHI and Sykdomspulsen.

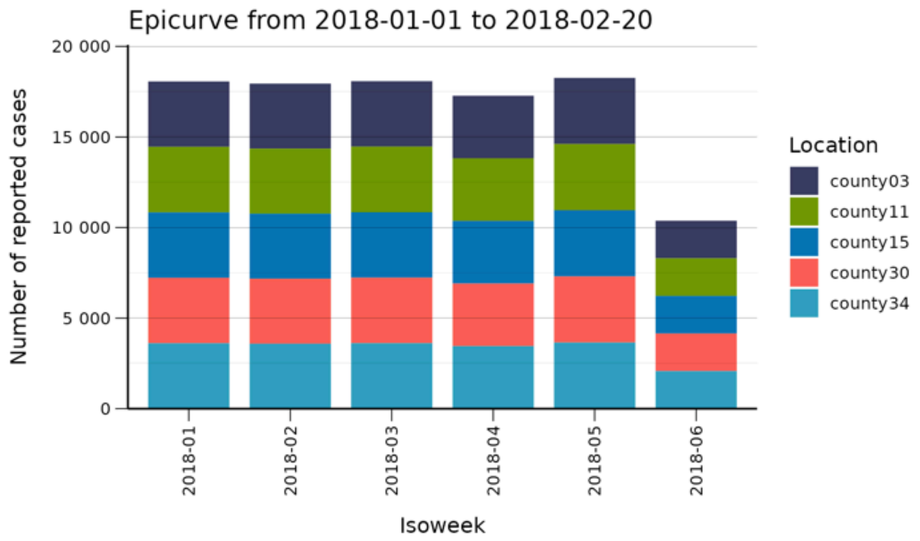
Read the introduction vignette here or run `help(package="org")`.

splstyle 2022.5.31 Get started Reference Articles Changelog

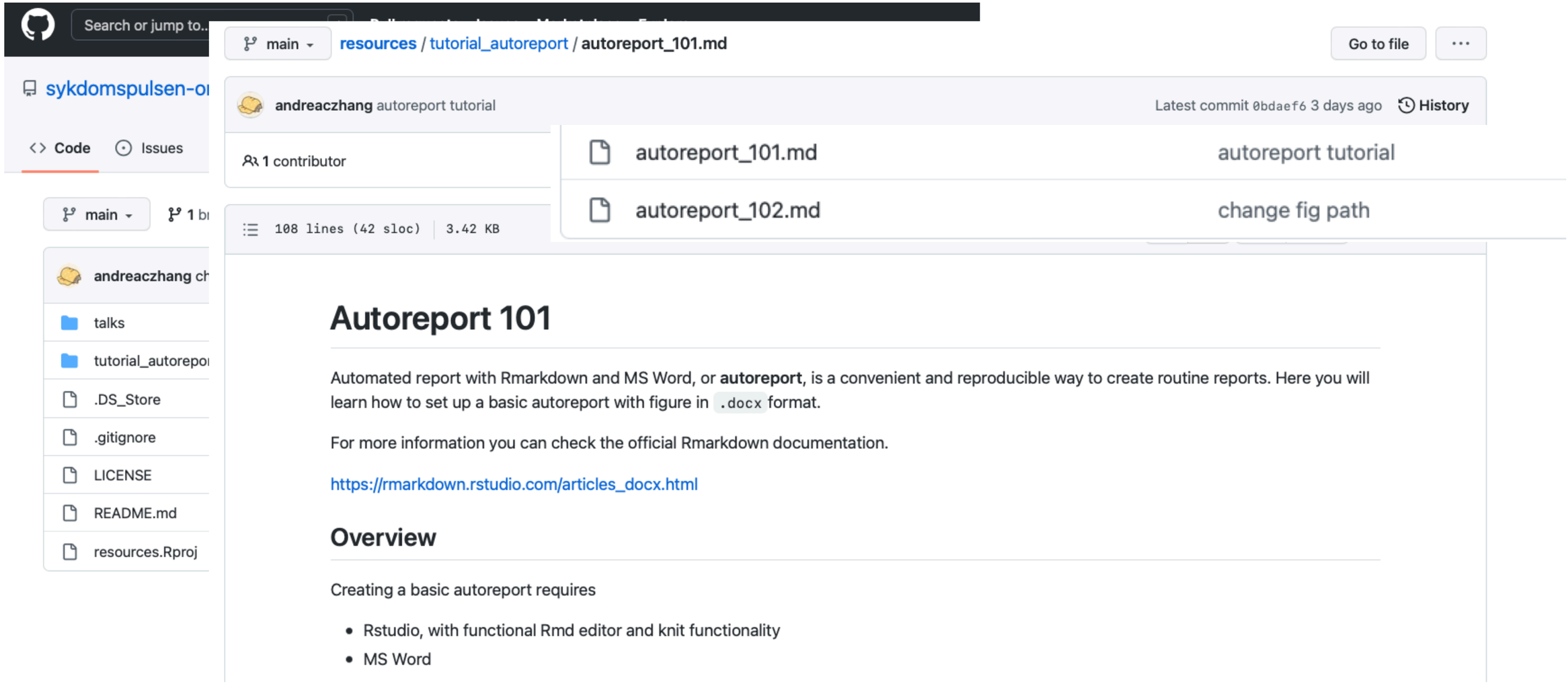
Coloured epicurves

Sometimes you would like to add colours to differentiate between different variables. This can be done through the `fill` attribute.

```
q <- ggplot(w[location_code %in% c(
  "county03",
  "county11",
  "county15",
  "county30",
  "county34")],
  aes(x = isoyearweek, y = N, fill = location_code))
q <- q + geom_col(width = 0.8)
q <- q + splstyle::scale_fill_fhi("Location",palette="primary")
q <- q + scale_x_discrete("Isoweek")
q <- q + scale_y_continuous("Number of reported cases",
  breaks = splstyle::pretty_breaks(5),
  expand = expansion(mult = c(0, 0.1))
)
q <- q + labs(title = "Epicurve from 2018-01-01 to 2018-02-20")
q <- q + labs(caption = fhi_caption())
q <- q + splstyle::theme_fhi_lines_horizontal()
q <- q + splstyle::set_x_axis_vertical()
q
```

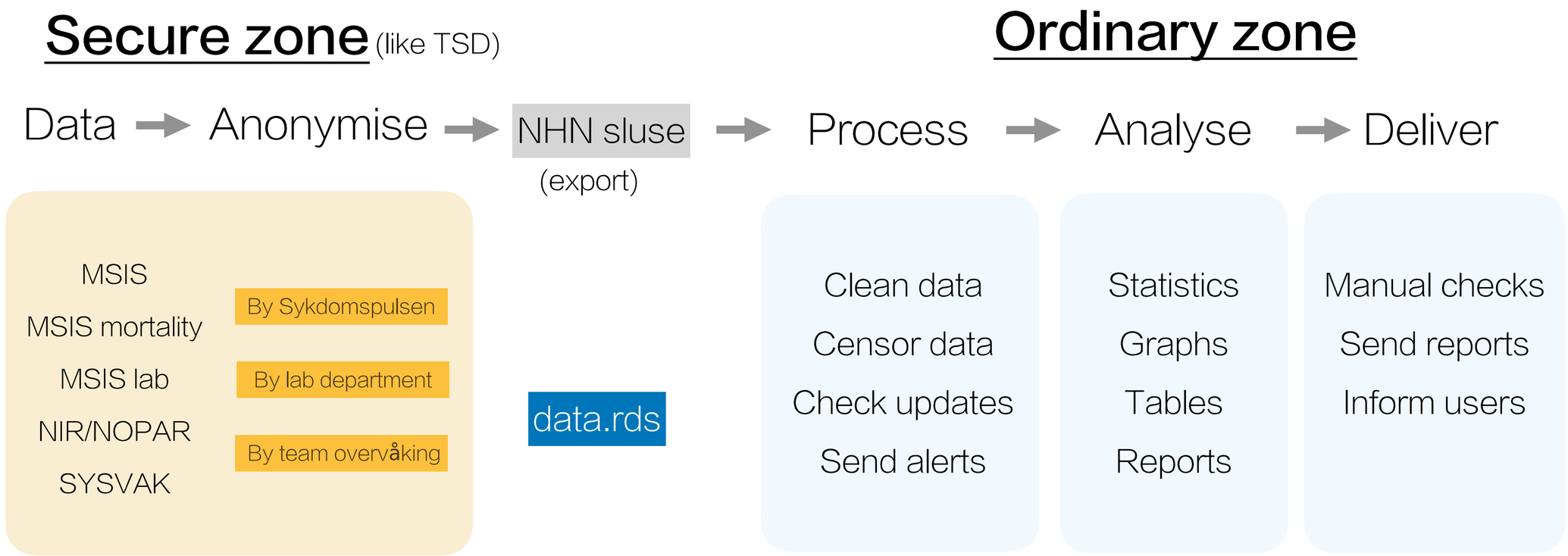


Bonus 2: autoreport tutorial



SP in production

From data to report



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

Scaling up

plnr for many analyses

Plan

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

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

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

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

...
```

Develop code for one plan, run for all plans

Convenient in parallel

Data pull

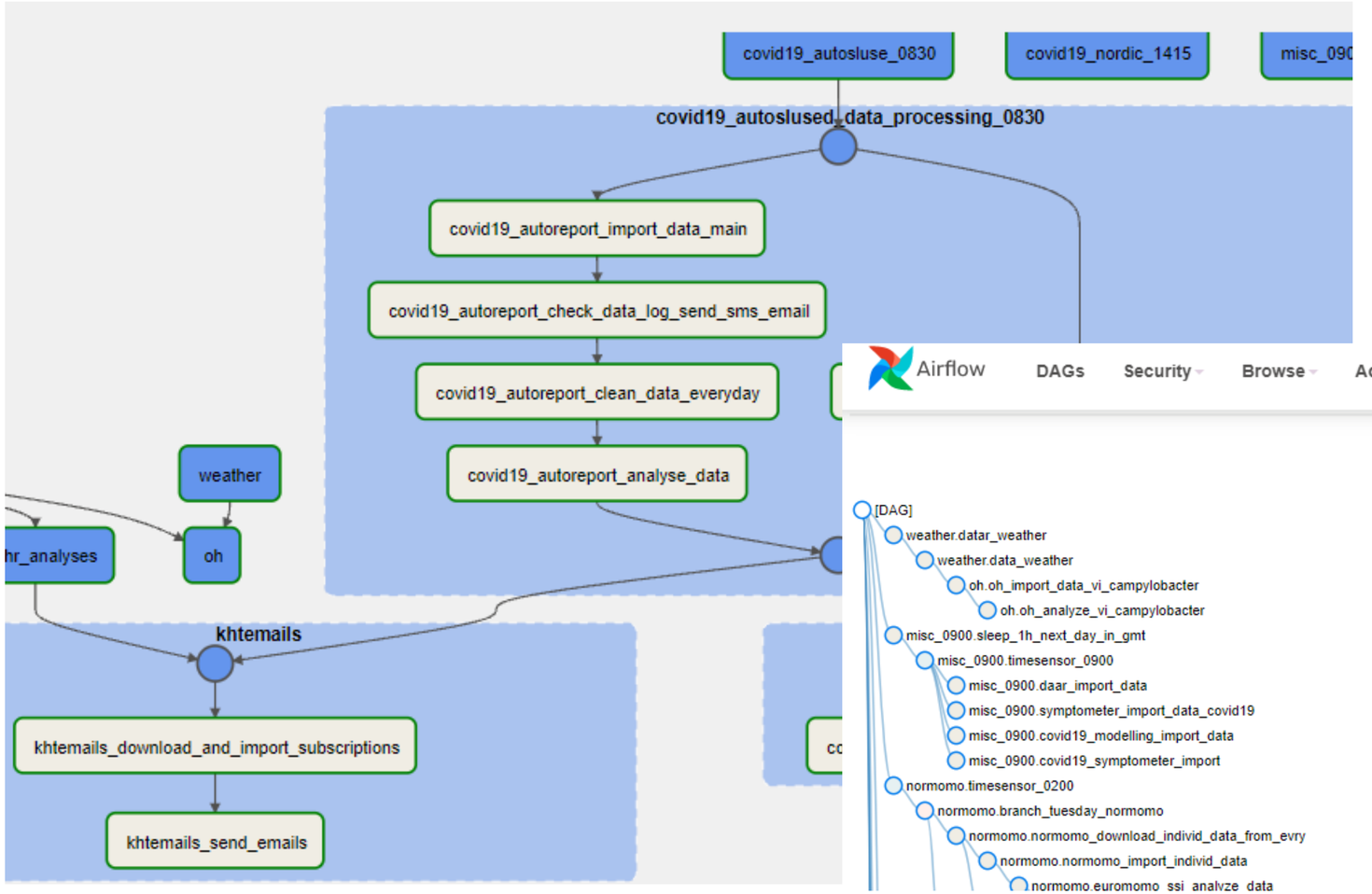
~800 dbtables

MSSQL + R data.table

```
field_types = c(
  "granularity_time" = "TEXT",
  "granularity_geo" = "TEXT",
  "country_iso3" = "TEXT",
  "location_code" = "TEXT",
  "border" = "INTEGER",
  "age" = "TEXT",
  "sex" = "TEXT",
  "date" = "DATE",
  "isoyear" = "INTEGER",
  "isoweek" = "INTEGER",
  "isoyearweek" = "TEXT",
  "season" = "TEXT",
  "seasonweek" = "DOUBLE",
  "calyear" = "INTEGER",
  "calmonth" = "INTEGER",
  "calyearmonth" = "TEXT",
  "tag_outcome" = "TEXT",
  "consultations_n" = "INTEGER",
  "denom_n" = "INTEGER",
  "consultations_pr100" = "DOUBLE",
  "consultations_baseline_n_expected" = "DOUBLE",
  "consultations_baseline_n_predinterval_l4" = "DOUBLE",
  "consultations_baseline_n_predinterval_l2" = "DOUBLE",
  "consultations_baseline_n_predinterval_u2" = "DOUBLE",
  "consultations_baseline_n_predinterval_u4" = "DOUBLE",
  "consultations_zscore" = "DOUBLE",
  "consultations_status" = "TEXT"
```

Automation

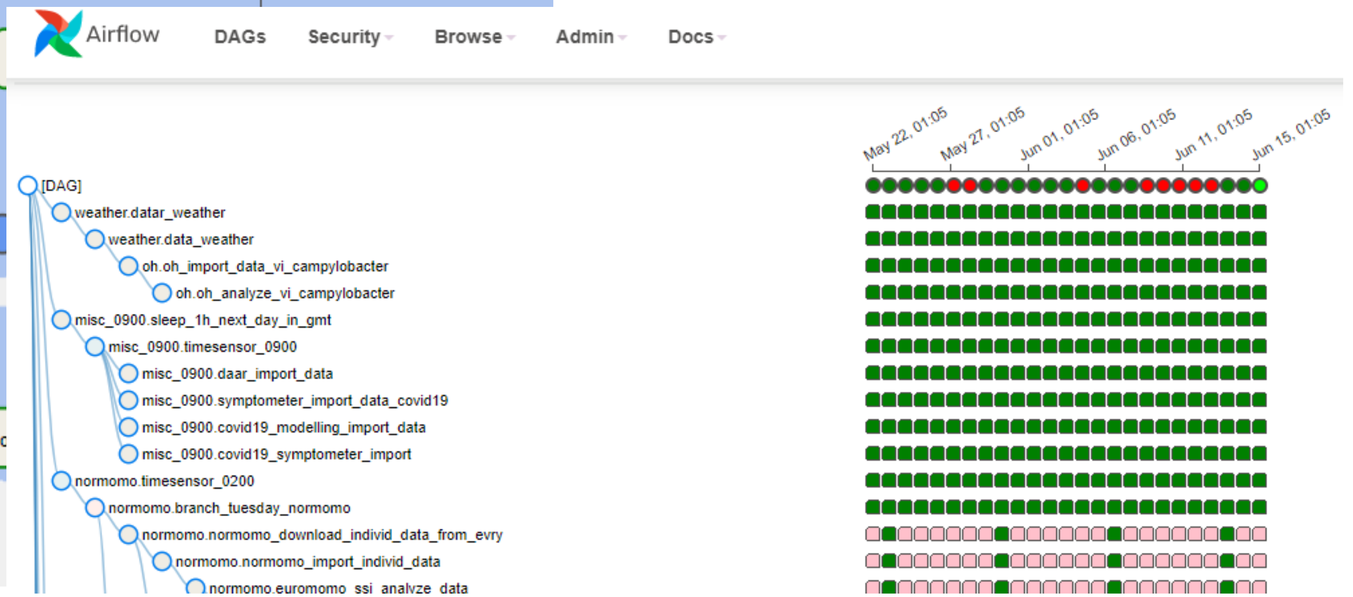
Airflow



We run 150+ tasks in airflow

Alert system: Email + SMS

Quality assurance: vakt
(debug, check report output)

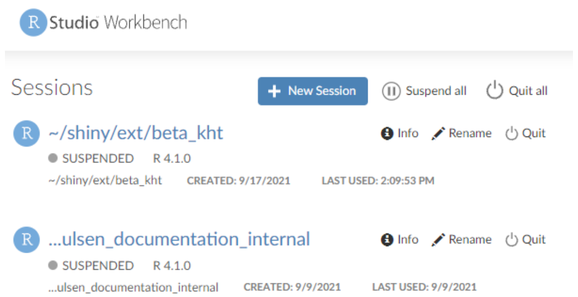


Development

RStudio, Github, Docker,CI/CD

Rstudio workbench

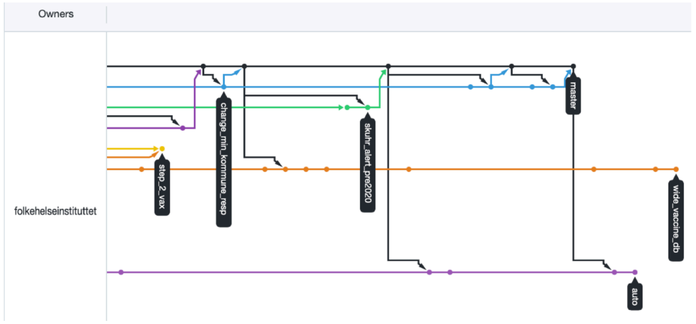
(Formerly Rstudio Server Pro)



100 USD per person per month

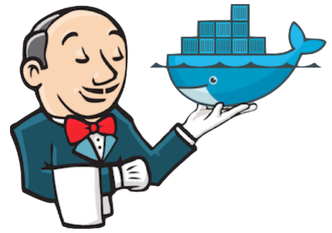
Github

One edit per branch



GoCD

Build docker image *

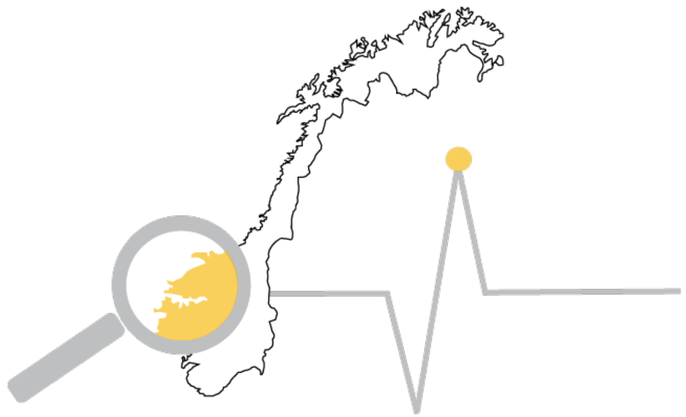


Contact our engineer for more details!

TL;DR

TL;DR

Sykdomspulsen



Automated real time public health surveillance platform

Open source

Efficient and scalable

Developed and used by statisticians

Learn more at https://docs.sykdomspulsen.no

GitHub repositories https://github.com/sykdomspulsen-org

Send us an email: sykdomspulsen@fhi.no

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/

Tidyverse principles

Tidyverse design principles https://design.tidyverse.org/unifying-principles.html

Cognitive speed: How the Tidyverse helped the British Red Cross respond quickly to COVID-19 (Rstudio::Global2021) https://www.rstudio.com/resources/rstudioglobal-2021/cognitive-speed-how-the-tidyverse-helped-the-british-red-cross-respond-quickly-to-covid-19/

Automated reports with Rmarkdown

Sykdomspulsen tutorial on autoreport https://github.com/sykdomspulsen-org/resources/blob/main/tutorial_autoreport/autoreport_101.md

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