

Scope of the possible with R

Brendan Clarke, NHS Education for Scotland, brendan.clarke2@nhs.scot

24/06/2024

Welcome

- this session is a **non-technical overview** designed for service leads
- we'll get going properly at 15.05
- if you can't access the chat, you might need to join our Teams channel:

tinyurl.com/kindnetwork

The KIND network

- a social learning space for staff working with knowledge, information, and data across health, social care, and housing in Scotland
- we offer social support, free training, mentoring, community events, ...
- Teams channel / mailing list

Session outline

- Why R, and why this session?
- R demo - take some data, load, tidy, analyse
- Strengths and weaknesses
 - obvious
 - less obvious
- Alternatives
- Skill development

R

- free and open-source
- multi-platform
- large user base
- prominent in health, industry, biosciences

Why this session?

- R can be confusing
 - it's code-based, and most of us don't have much code experience
 - it's used for some inherently complicated tasks
 - it's a big product with lots of add-ons and oddities
- But R is probably the best general-purpose toolbox we have for data work at present
 - big user base in health and social care
 - focus on health and care-like applications
 - not *that* hard to learn
 - extensible and flexible
 - capable of enterprise-y, fancy uses

R demo

- this is about showing what's possible, and give you a flavour of how R works
- we won't explain code in detail during this session
- using **live open data**

Weekly A&E Activity and Waiting Times

This is the new weekly Accident & Emergency (A&E) open data area. From 2 May 2023 the weekly attendance and waiting times statistics at Emergency Departments (EDs) across Scotland will be published here. The monthly A&E open data statistics can be found [here](#).

This release by Public Health Scotland (PHS), provides a weekly update of key statistics on attendances at Emergency Departments (EDs) across Scotland. The information includes trends in the number of attendances and proportion waiting over 4, 8 and 12 hours.

These statistics do not reflect all types of A&E activity and they should not be used to monitor against the 4-hour A&E access standard.

Further information on A&E services across Scotland can be found [here](#).

Load that data

```
1 ae_activity <- read_csv("data/weekly_ae_activity_20240609.csv")
```

One small bit of cheating: renaming

```
1 names(ae_activity) <- c("date", "country", "hb", "loc", "type", "attend", "n_within", "n_4", "perc_4", "n_8")
```

Preview

Preview of data

date	country	hb	loc	type	attend	n_within	n_4	perc_4	n
20220731	S92000003	S08000031	G107H	Emergency Department	1582	972	610	61.4	1
20230730	S92000003	S08000022	H103H	Emergency Department	167	156	11	93.4	
20240519	S92000003	S08000030	T101H	Emergency Department	1316	1121	195	85.2	
20181223	S92000003	S08000032	L308H	Emergency Department	1338	1245	93	93.0	

date	country	hb	loc	type	attend	n_within	n_4	perc_4	n
20190224	S92000003	S08000029	F704H	Emergency Department	1335	1164	171	87.2	

Removing data

```
1 ae_activity <- ae_activity |>
2   select(!c(country, contains("perc_")))
```

Preview of data

date	hb	loc	type	attend	n_within	n_4	n_8	n_12
20230604	S08000020	N121H	Emergency Department	369	326	43	0	0
20220410	S08000031	G513H	Emergency Department	1134	1117	17	0	0
20180304	S08000016	B120H	Emergency Department	399	372	27	2	1

Tidying data

```
1 ae_activity <- ae_activity |>
2   mutate(date = ymd(date))
```

Preview of data

date	hb	loc	type	attend	n_within	n_4	n_8	n_12
2018-01-28	S08000031	C418H	Emergency Department	1228	1098	130	10	0
2016-11-06	S08000020	N101H	Emergency Department	1034	968	66	2	1
2016-11-13	S08000024	S308H	Emergency Department	1023	992	31	0	0

Subset data

- we'll take a random selection of 5 health boards to keep things tidy

```
1 ae_activity <- ae_activity |>
2   filter(hb %in% boards)
```

Preview of data

date	hb	loc	type	attend	n_within	n_4	n_8	n_12
2019-10-13	S08000031	G513H	Emergency Department	1355	1277	78	0	0
2017-08-27	S08000026	Z102H	Emergency Department	142	139	3	0	0

Basic plots

```
1 ae_activity |>
2   ggplot() +
3   geom_line(aes(x = date, y = attend, colour = hb, group = loc))
```



Joining data

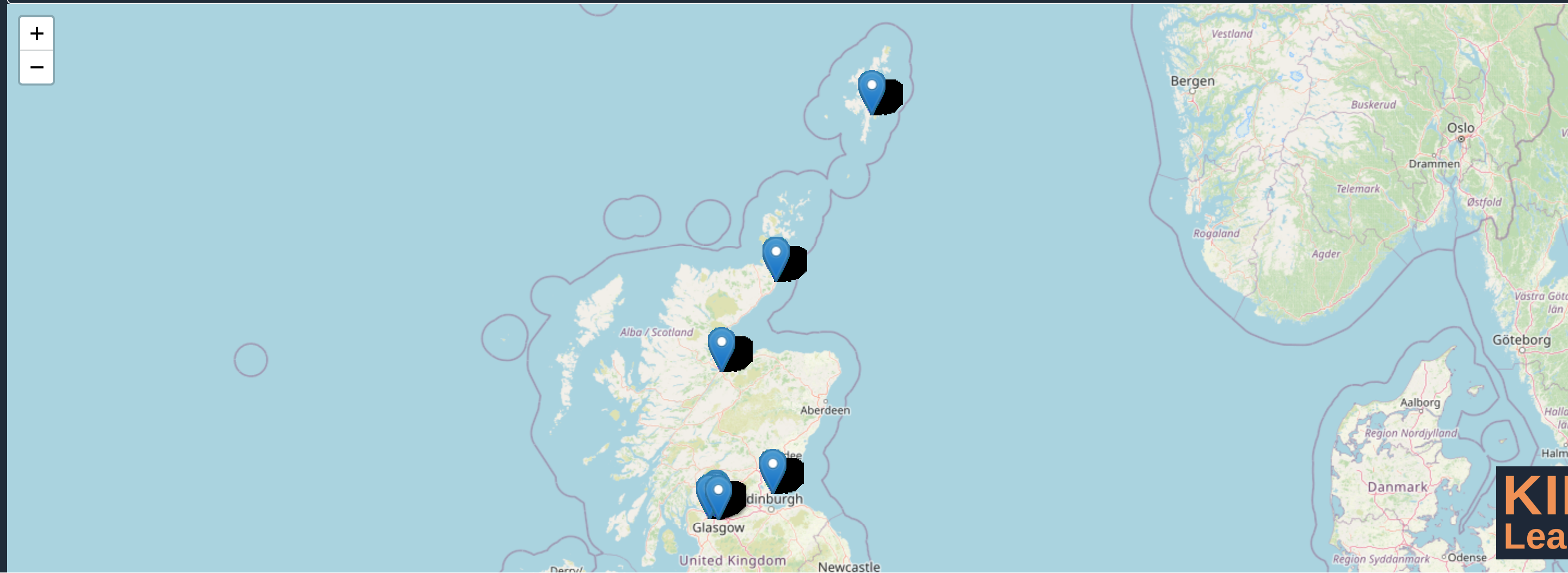
```
1 ae_activity |>
2   left_join(read_csv("data/boards_data.csv"), by = c("hb" = "HB")) |>
3   select(!any_of(c("_id", "HB", "HBDateEnacted", "HBDateArchived", "Country"))) |>
4   ggplot() +
5   geom_line(aes(x = date, y = attend, colour = HBName, group = loc))
```



and again...

Add to a map

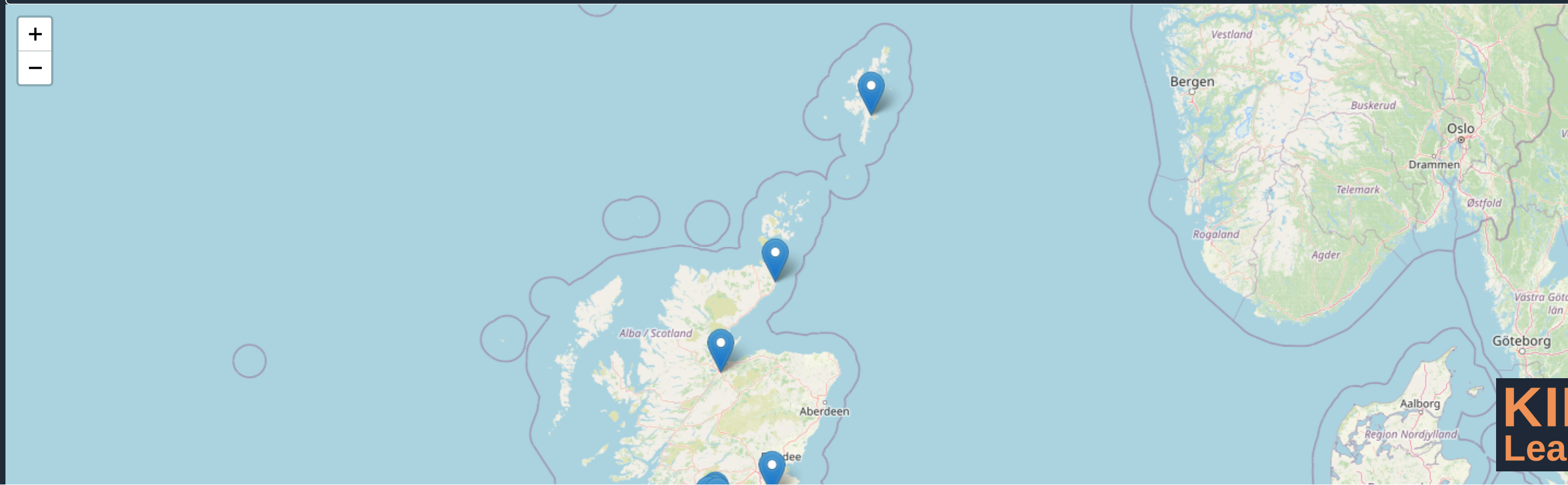
```
1 ae_activity_loc |>  
2   leaflet::leaflet() |>  
3   leaflet::addTiles() |>  
4   leaflet::addMarkers(~longitude, ~latitude, label = ~HospitalName)
```

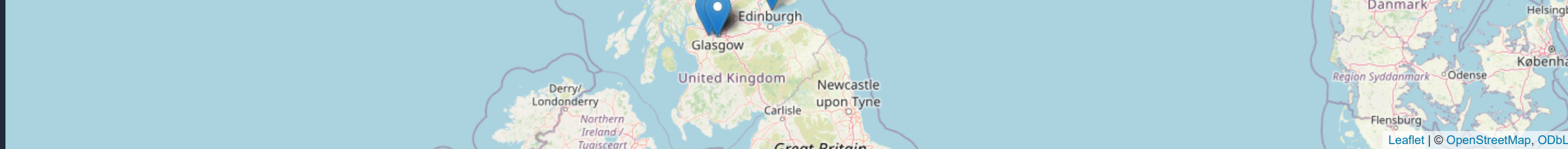




Then make that map more useful

```
1 ae_activity_loc |>
2   group_by(HospitalName) |>
3   summarise(attend = sum(attend), n_within = sum(n_within), longitude = min(longitude), latitude = min(latitude)) |>
4   mutate(rate = paste(HospitalName, "averages", scales::percent(round(n_within / attend, 1)))) |>
5   leaflet::leaflet() |>
6   leaflet::addTiles() |>
7   leaflet::addMarkers(~longitude, ~latitude, label = ~rate)
```





Then add to reports, dashboards...

Strengths

- enormous scope and flexibility
- a force-multiplier for fancier data work
 - helps collaboration within teams, between teams, between orgs
 - reproducible analytics
 - modular approaches to large projects
- decreasing pain curve: the fancier the project, the better

Weaknesses

- harder to learn than competitors
- very patchy expertise across H+SC Scotland
- complex IG landscape
- messy skills development journey

Skill development

Session	Date	Area	Level
Iteration in R	09:30-11:00 Fri 5th July 2024	R	🔪🔪 : intermediate-level
Getting more out of dplyr	10:30-12:00 Wed 17th July 2024	R	🔪🔪 : intermediate-level
Testing R code	15:00-16:30 Wed 7th August 2024	R	🔪🔪 : intermediate-level

Chat, queries, questions

- Teams channel: tinyurl.com/kindnetwork
- weekly email update: tinyurl.com/kindupdate
- get in touch: brendan.clarke2@nhs.scot
- session feedback link at resource page