UNIVERSITY OF OSLO

OSLO UNIVERSITY HOSPITAL

Explore Real-World hospital Electronic Health Records data with ggehr

Chi Zhang, PhD

OCBE (Oslo Center for Biostatistics and Epidemiology)
@ University of Oslo & Oslo University Hospital

R in Pharma, 24.10.31

Website: https://andreaczhang.github.io/
LinkedIn: Chi Zhang
Email: chizapoth@gmail.com

About me

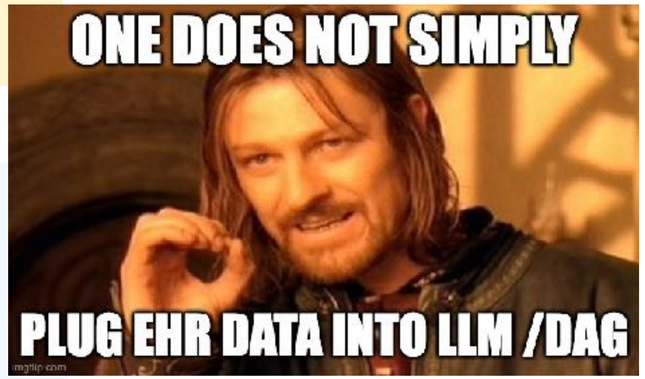
- Statistician - RWD: **hospital EHR and large public health registries** (PhD on EHR)
- R developer, used **R packages and Shiny** for large-scale public health surveillance in Norway; **Quarto, webR** for teaching
- Co-lead of **CAMIS** (PHUSE DVOST project), contributor to **RWD guideline** for statistical programmer

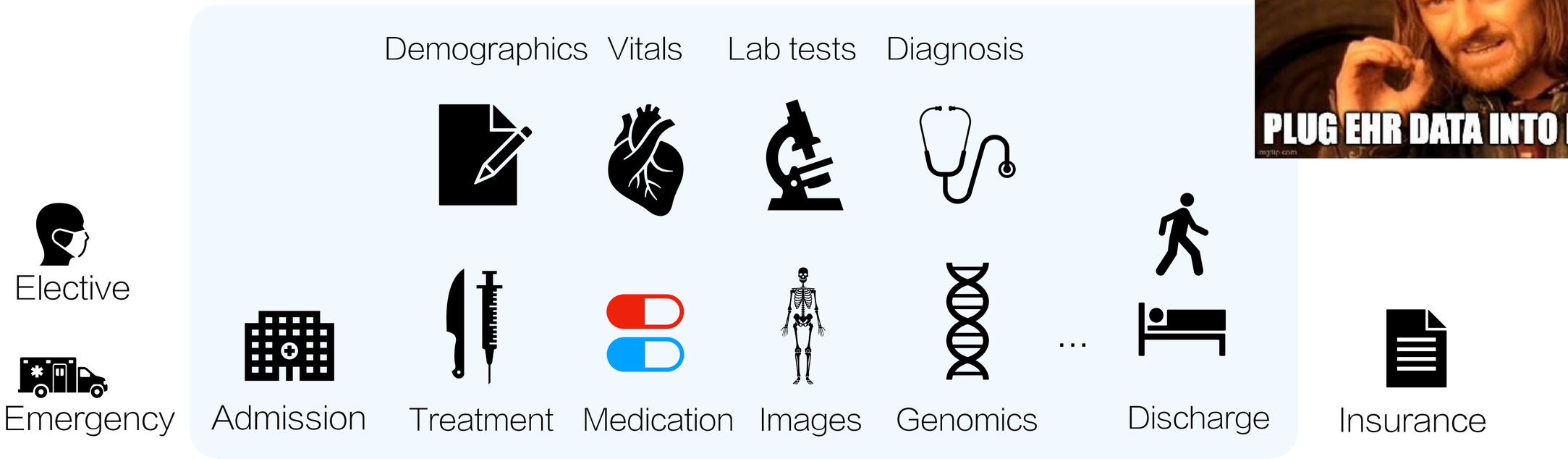
Disclaimer: opinions in this talk are all mine and do not reflect that of my employer

Electronic Health Records

EHR data is generated through out the patient encounter with the healthcare system.

Not research quality data; not RCT, observational in nature; multi-modal, mixed type, temporal x static; errors and missing …





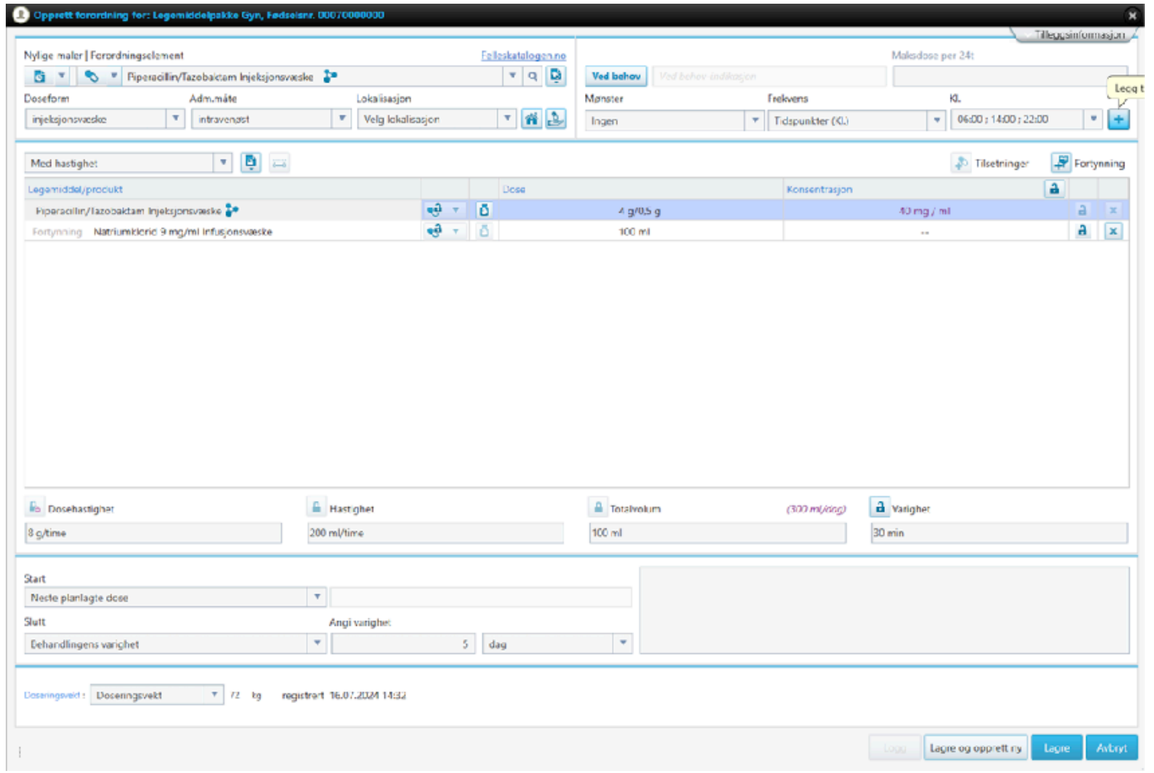
EHR system example: MetaVision

MetaVision from iMD-soft

Used by **27 countries, 400+ hospitals, 2M patients** per year (e.g. NHS in UK, Univeristy Hospital Geneva)

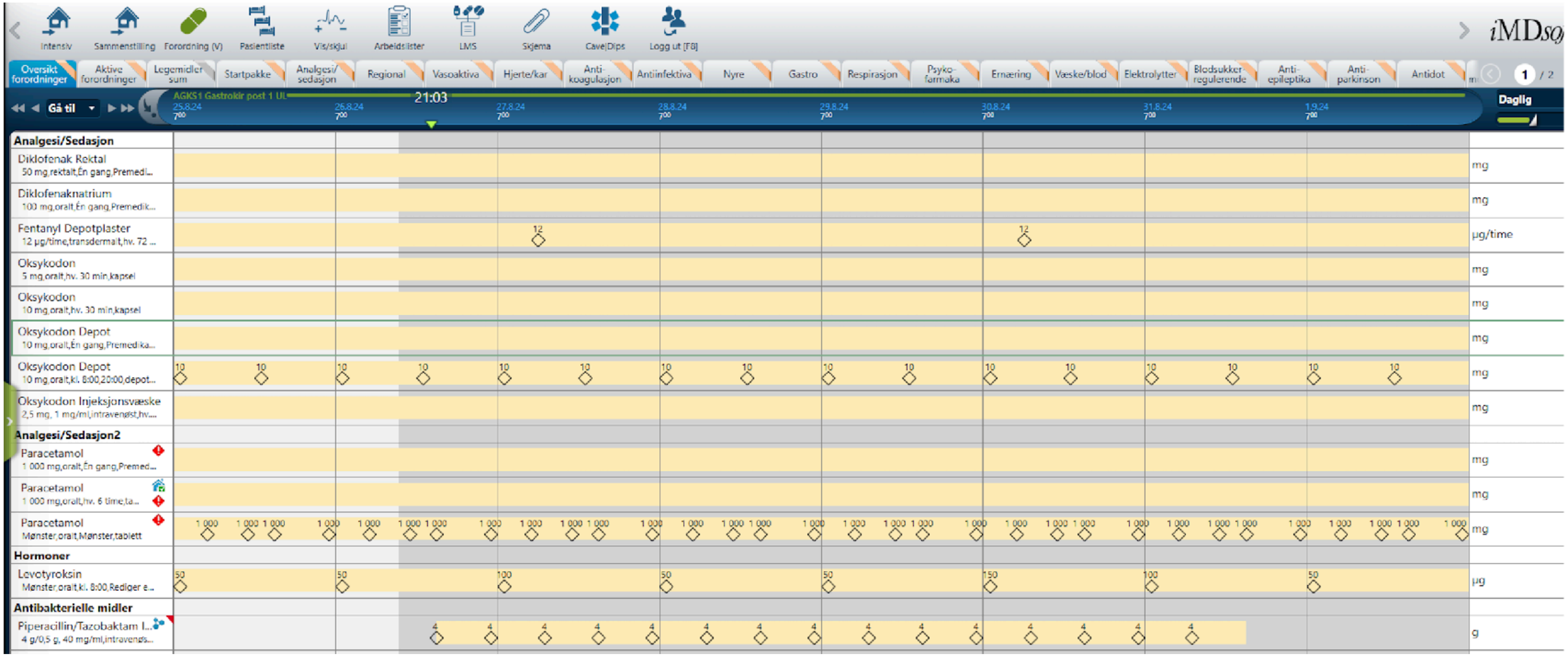
Point and click, drop-down menus to select medication, pre-configured for drug, dose, duration

Prescription and **use** are recorded separately



https://www.imd-softde.de/

Medication records in MV (pseudo patient data)



What data analysts see

B	C	D	E	F	G	H
KontaktID	AB_SkjemaTidRelativ(Timer)	AntibiotikaSkjema	AB_ForsteGang	Profylakse	Samfunnservervet	Sykehuservervet
11439	-30	Ampicillin	Ja		Akutt buk/peritonitt	
11439	-30	Gentamicin	Ja		Akutt buk/peritonitt	
11439	-30	Metronidazol	Ja		Akutt buk/peritonitt	
11439	-24	Ampicillin			Akutt buk/peritonitt	
16406	2607	Cefalotin	Ja	Perioperativ kirurgisk		
16406	2609	Cefalotin	Nei	Perioperativ kirurgisk		
16406	2612	Cefalotin	Nei	Perioperativ kirurgisk		
13549	829	Klindamycin	Ja	Perioperativ kirurgisk		
13549	830	Klindamycin	Nei	Perioperativ kirurgisk		
12577	2605	Cefalotin		Perioperativ kirurgisk		
12577	2606	Cefalotin	Nei	Perioperativ kirurgisk		
10245	6858	Metronidazol		Perioperativ kirurgisk		
11341	5194	Cefalotin		Perioperativ kirurgisk		
11341	5196	Cefalotin	Nei	Perioperativ kirurgisk		
338	4940	Pivmecillinam	Nei		Cystitt	
17279	2561	Kloksacillin	Nei		Infeksjon i ben og ledd	
17279	2563	Kloksacillin	Ja	Perioperativ kirurgisk		
13722	2257	Piperacillin/Tazobaktam			Akutt buk/peritonitt	
13722	2337	Ciprofloksacin			Akutt buk/peritonitt	
13722	2337	Metronidazol			Annen infeksjon	
13722	2337	Trimetoprim/Sulfametoksazol	Nei		Annen infeksjon	
13722	2361	Metronidazol	Nei		Annen infeksjon	
4144	5428	Cefalotin		Perioperativ kirurgisk		
4144	5433	Cefalotin	Ja	Perioperativ kirurgisk		
1808	5462	Fenoksymetylpenicillin	Ja		Pneumoni	
1808	5463	Benzylpenicillin	Ja			Pneumoni
13285	5364	Klindamycin	Ja	Perioperativ kirurgisk		

One excel sheets for all patients together

Many empty cells, not necessarily missing data

Potential errors in some columns

What data analysts see

A	B	C	D	E	F	G
Kontak	PID	Value	ParameterName	TidTilProsess	Forbruk	Benevnelse
10009	1	1.0000000	#NAME?	7978	1889.0000000000	Min
10009	1	1.0000000	#NAME?	7978	67.000000000000	Min
10009	1	1.0000000	- PVK (1,3 mm/18 G)	7978	88.000000000000	Min
10009	1	1.0000000	- PVK (1,3 mm/18 G)	7942	2040.000000000	Min
10009	1	1.0000000	- PVK (1,1 mm/20 G)	7976	2004.000000000	Min
10009	1	1.0000000	- PVK (1,1 mm/20 G)	7976	2004.000000000	Min
11439	3	0.0187500	Metronidazol Injeksjonsvæske	-29	1.5000000000000	Gram
11439	3	0.0250000	Metronidazol Injeksjonsvæske	-17	1.0000000000000	Gram
11439	3	0.0250000	Metronidazol Injeksjonsvæske	30	1.0000000000000	Gram
11439	3	0.0083333	Gentamicin Injeksjonsvæske	-6	0.5000000000000	Gram
11439	3	0.0666666	Ampicillin Injeksjonsvæske	17	2.0000000000000	Gram
11439	3	0.0666666	Ampicillin Injeksjonsvæske	23	2.0000000000000	Gram
11439	3	0.0666666	Ampicillin Injeksjonsvæske	30	2.0000000000000	Gram

When jointly analyzed with another table, some **misalignments of time**

3.4% of all records have medication usage *before* prescription

15% antibiotics use have **mismatches**: prescribed but not used; vice versa

What data analysts see

Physicians and infection control team usually have a decent idea of what is going on at the hospital, but analysts don’t!

Is there a quick way to understand **how events happen** around the patients?

How is the **data quality**?

What are the possible causes for **errors, mismatches and missingness**?

An R package for EHR visualization

ggplot2 extension for EHR data

gg E-H-R

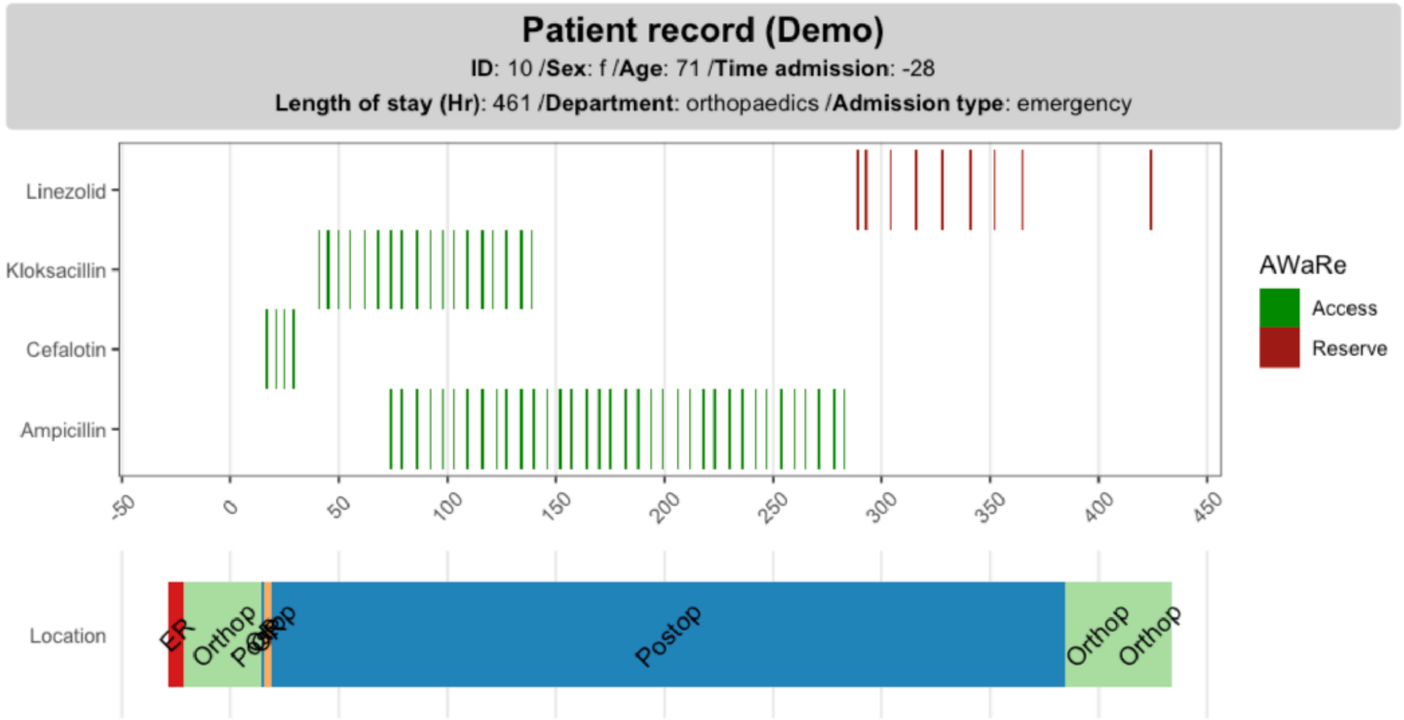
(Early stage, not ready for release, data simulated as close as possible to reality)

A set of R tools to visualize **mixed type of information** related to a patient

Motivations:

Overview of events happened to one patient, **ordered by time**

Identify errors / strange recordings



Multiple layers of information

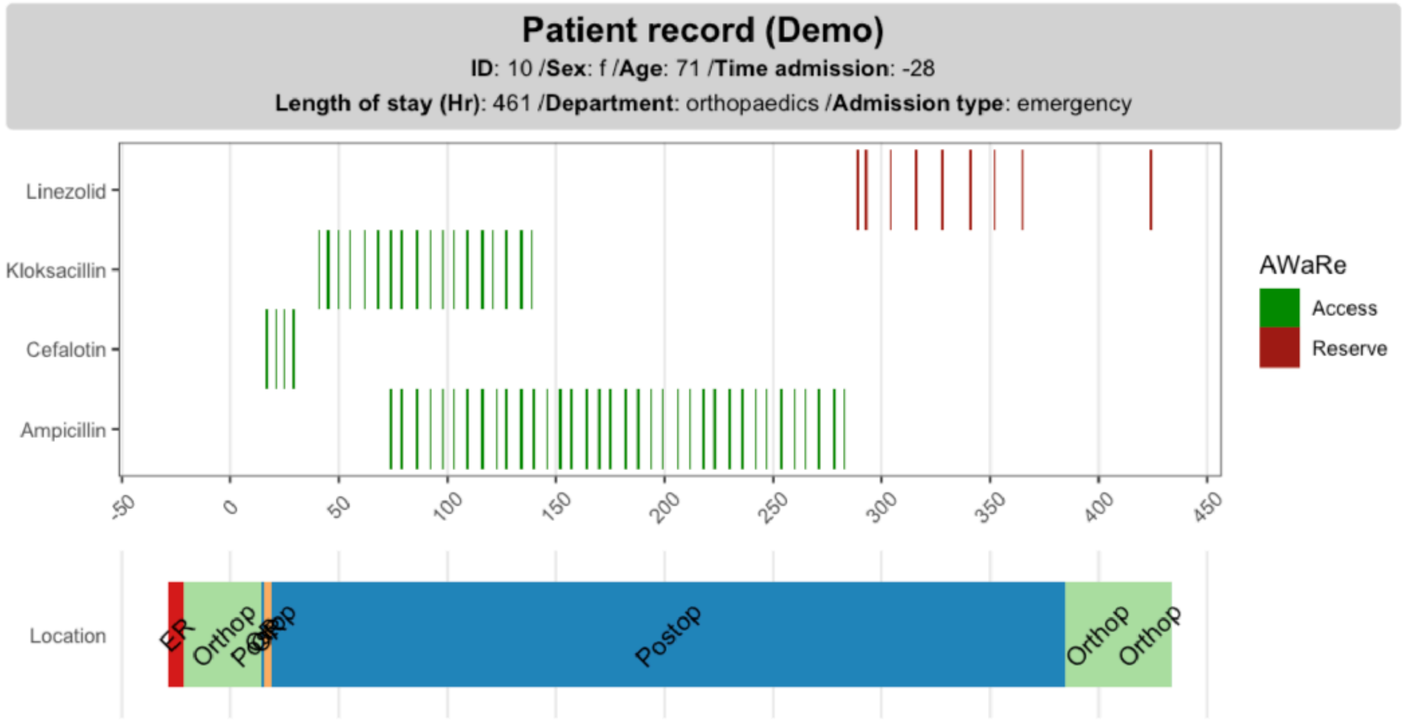
Mixed data types

Static data: demographics, admission information

Temporal: treatments (drug, procedure), locations

Layers

Can add more layers, such as **locations**, physiological time series (**temperature**, **NEWS score**) and **catheters**, etc



Drug prescription and usage

Drug prescription and use plotted together

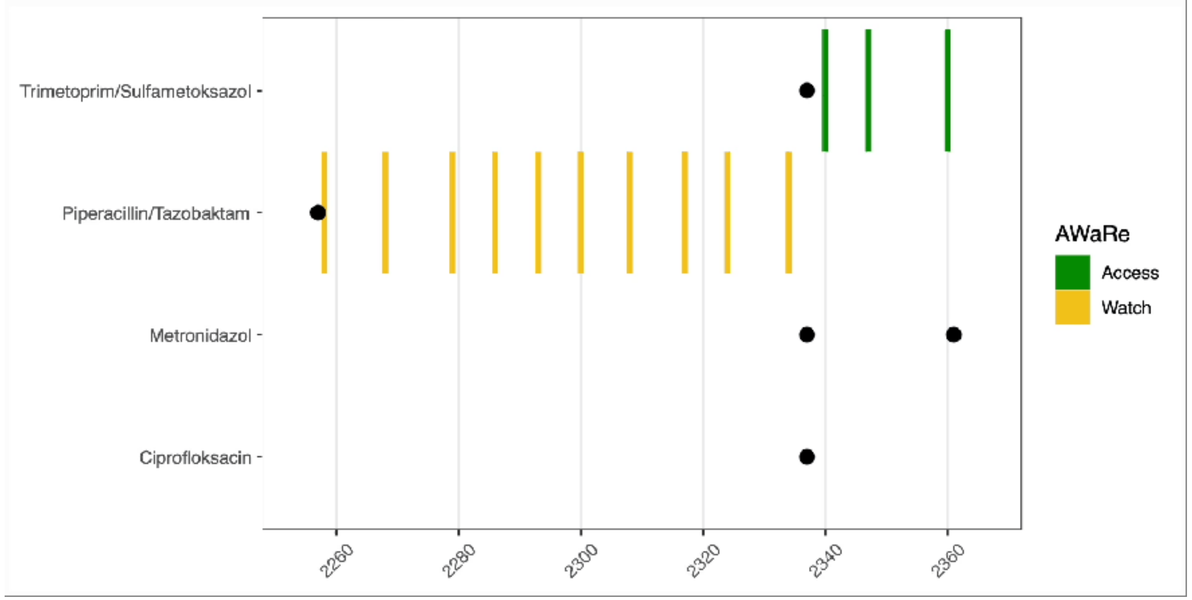
Drugs (e.g. antibiotics) colored based on **WHO AWaRe** categories

Can identify if the use of drug has **escalated**

Visually **identify if errors/strange things** exist in data recording

Can visualize prescription time and purpose, **interactively**

```
p1i <- plot_drug_prescribe_interactive(data_prescribe = dpatient_presc,
                                       plot_obj = p1p)
p1i
```



AWaRe: Access, Watch, Reserve - unrestricted use, use with care, use as last resort

Patient background card

```
# demographics ----#
dpatient_demoinfo <- make_demographic_info(demo_df = dpatient_demog)
dpatient_demoinfo
#> [1] "<b>Patient record (Demo)</b><br><span style = 'font-size:10pt'>**ID**: 1 ,
```

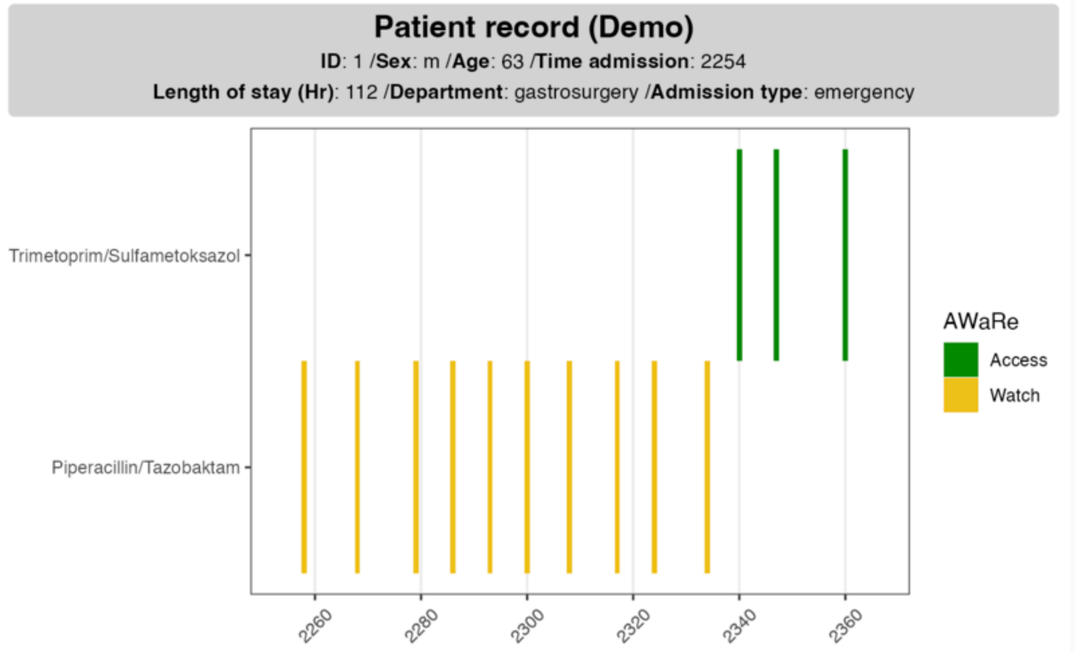
Static information card provides patient background information

Age, sex, admission related (department, type)

Potentially also diagnosis

Relevant when discussing with clinicians to understand what kind of patient he/she is

```
p1pp <- plot_info_demographic(info_text = dpatient_demoinfo,
                              plot_obj = p1)
p1pp
```

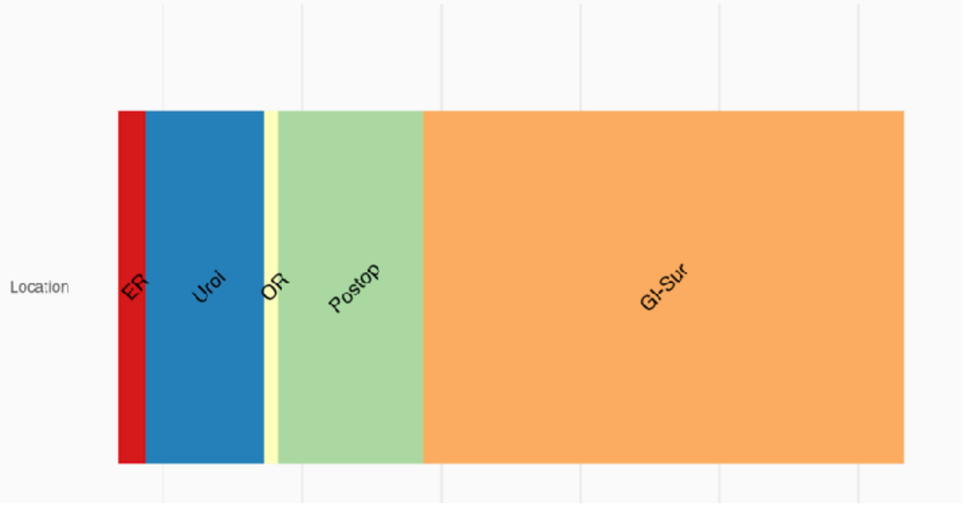


Other information

Information such as **location** can also be visualized as **color blocks**

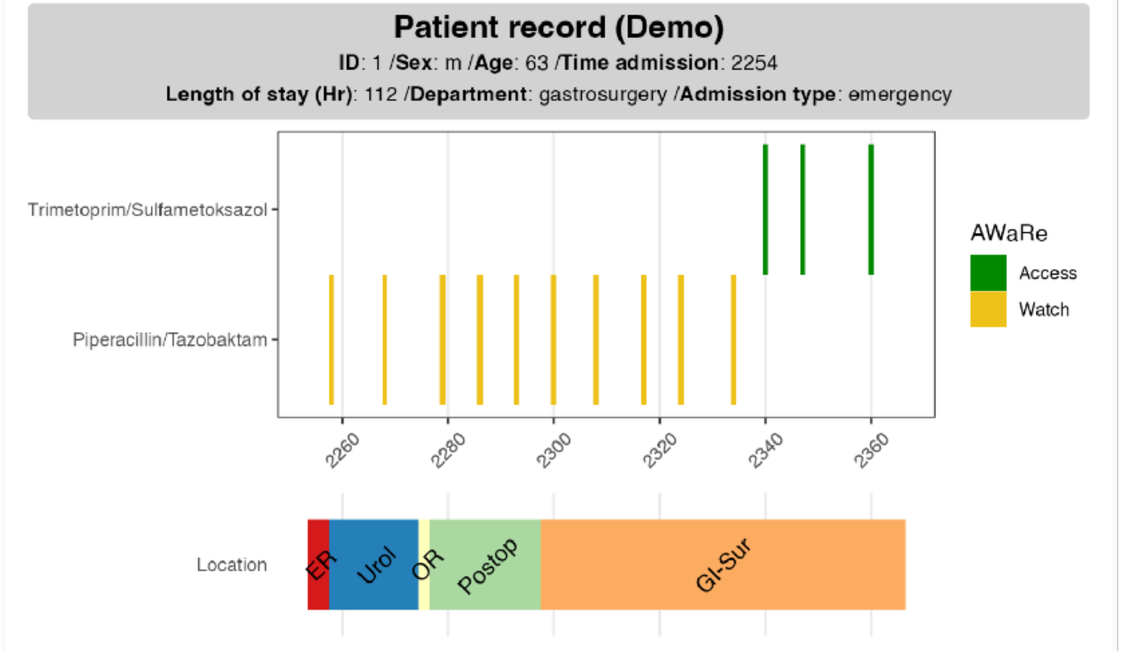
```
# location ----#
dloc <- make_location(data_location = dpatient_loc,
                      tadmin = dpatient_demog$t0,
                      los = dpatient_demog$los)
#> Joining with `by = join_by(location_code)`
#> Joining with `by = join_by(location_code)`

p2 <- plot_location(loc_obj = dloc, keep_time = F)
p2
```



Multiple layers stacked with **`patchwork`**

```
p1pp + p2 + plot_layout(ncol = 1, heights = 2:1)
```



ggplot2 extension for EHR data

Why not shiny or quarto?

- limited resources, maintenance
- **not yet production level**
- Rpkg can be used by everyone

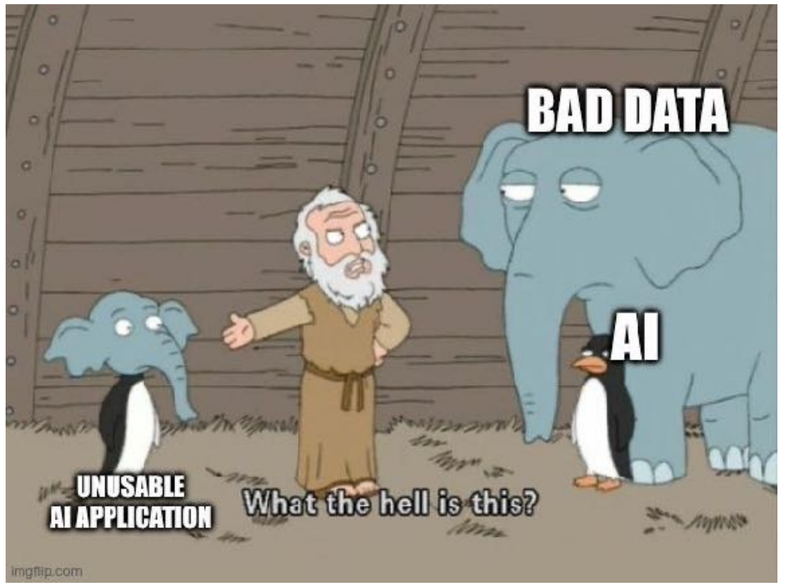
Better EDA -> better research questions

Early stage of development

If you have a **use-case** and/or **dataset** to demonstrate with the package, welcome to create an issue or talk to me!

GitHub: **`andreaczhang/ggehr`**

Next step: aggregated **visualization for a cohort** -> assist causal hypothesis generation



Looking for new opportunities as data scientist / R dev / statistician. Let's chat!