

Extracting and deploying intelligence from text-based patient experience data

An open source project using Python and R

Bristol Data Science Seminar Series
20 May 2021

Andreas Soteriades - Data Scientist
Milan Wiedemann - Data Scientist

Clinical Development Unit, Nottinghamshire Healthcare NHS Foundation Trust

WHO ARE WE

WHAT WE ARE DOING



Chris Beeley, PhD

Senior Analyst 🧙

Background: Clinical Psychology

Owens two cats named Pavlova and Freud

Shiny, server hosting



Lori Edwards Suárez

Clinical Analyst 🦄

Background: NHS Expert

RiO Expert, Data Analysis



Andreas Soteriades, PhD

Data Scientist 🧠

Background: Applied Mathematics

Knows more about cows than anyone else in this room

Machine Learning

Zoë Turner

Senior Information Analyst 🧑🍳

Background: Philosophy

Knows too much about data

SQL and data engineering



Milan Wiedemann, PhD

Data Scientist 🐛

Background: Clinical Psychology

Has no memory for NHS acronyms

Data Analysis

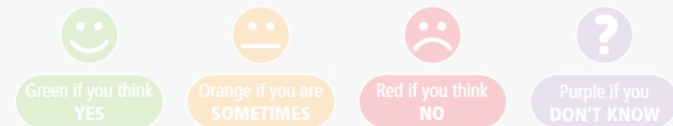


Please answer all questions from your perspective and about your own experiences.

1. Overall, how was your experience of our service?

Very good Good OK Bad Very bad Don't know

For each question, please pick:



2. Were our services good at listening to you?



3. Were our services good at explaining things clearly?



4. Were our services good at being kind to you?



5. Were our services good at involving you in the way we helped the person you care for?



6. Were our services good at talking to you about the support you could have as a carer?



7. Did our services tell you who you could contact if you were worried about the person you care for?



8. Did we tell you who to contact for a carers' assessment?



9. What was good?

XXX has saved my life. Thank you xxx

10. What could we do better?

More freedom - to be trusted more.

☐ Please tick here if you **DO NOT** wish your comments to be made public

Would you prefer to give us feedback online?

- Complete this survey at bit.ly/NottsHCfeedback
- Post your story on CareOpinion.org.uk

Could you use your experience to help us improve our services? Contact the Involvement, Experience and Volunteering Team on: 0115 956 0845 or involve@nottshc.nhs.uk

Text Classification

More freedom - to be trusted more.

Privacy and Dignity

XXX has saved my life. Thank you xxx

Service Quality / Outcomes

Sentiment Analysis

Need more staff to take **pressure** off.

Negative

The service was **great**.

Positive

Machine learning
methods and interface

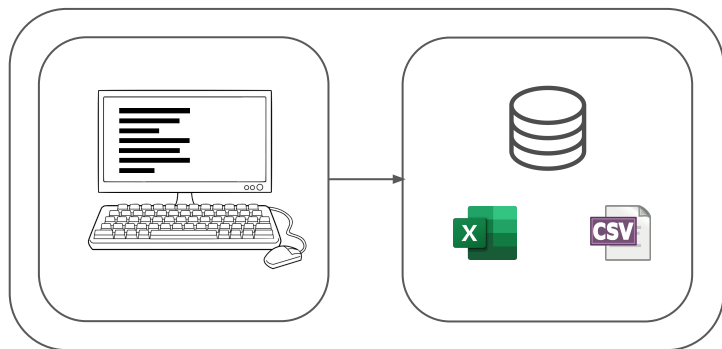
https://involve.nottshc.nhs.uk:8443/text_mining_dashboard/

Data exploration
methods and interface

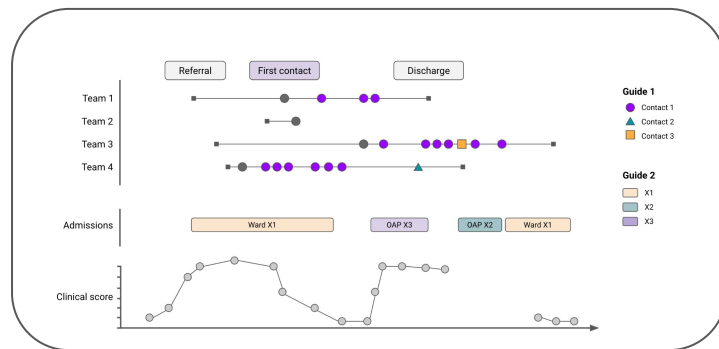
<https://involve.nottshc.nhs.uk:8443/experiencesdashboard/>

A GLIMPSE INTO
OUR WORK

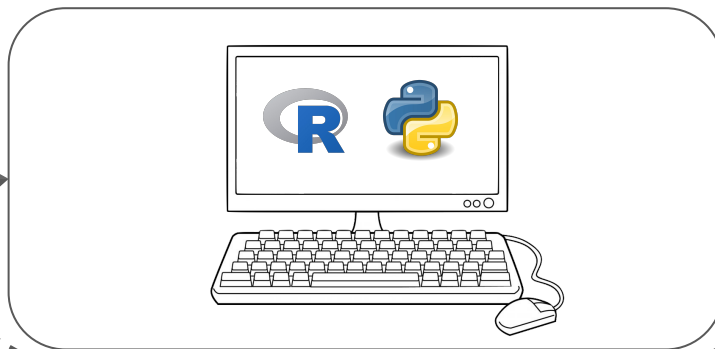
Data



Results



Understand data



Tidy

Analyse

Communicate

Visualise

BIG PROBLEM

UNINFORMED* SOLUTION

*usually the first idea that comes to mind or the first answer on Stack Overflow

SLIGHTLY* BETTER SOLUTION

*sometimes also much better solution

Q



Q

Skills

Python + {reticulate}
Git + GitHub
Testing

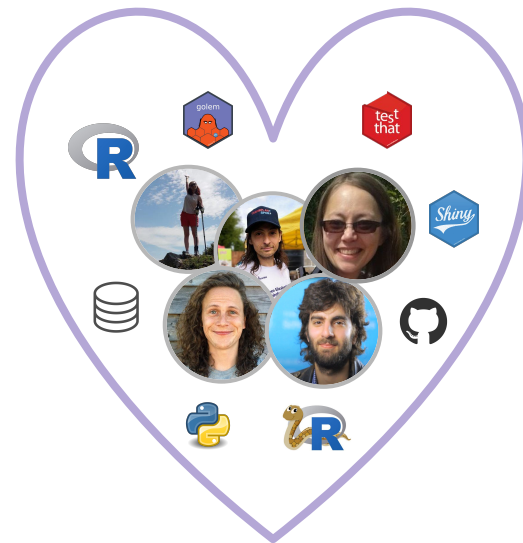
Code review
Live coding
Book club



Q

Productionising

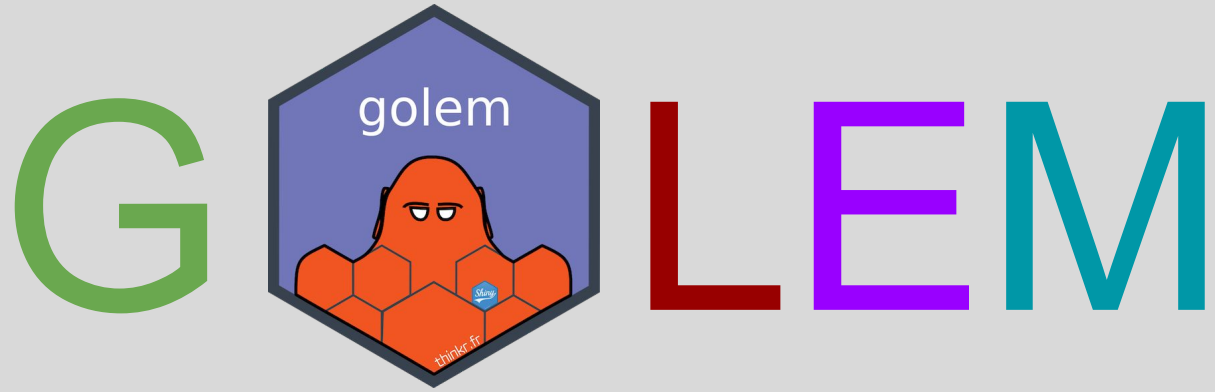
R/Python packages
Golem
Documentation
User feedback
Roll-out



Developing ideas

Developing as a team

Developing tools



Dashboards

**similar dashboard for different
NHS Trusts**

messy {shiny} code

{golem} + helper R pkgs + open source

Golem



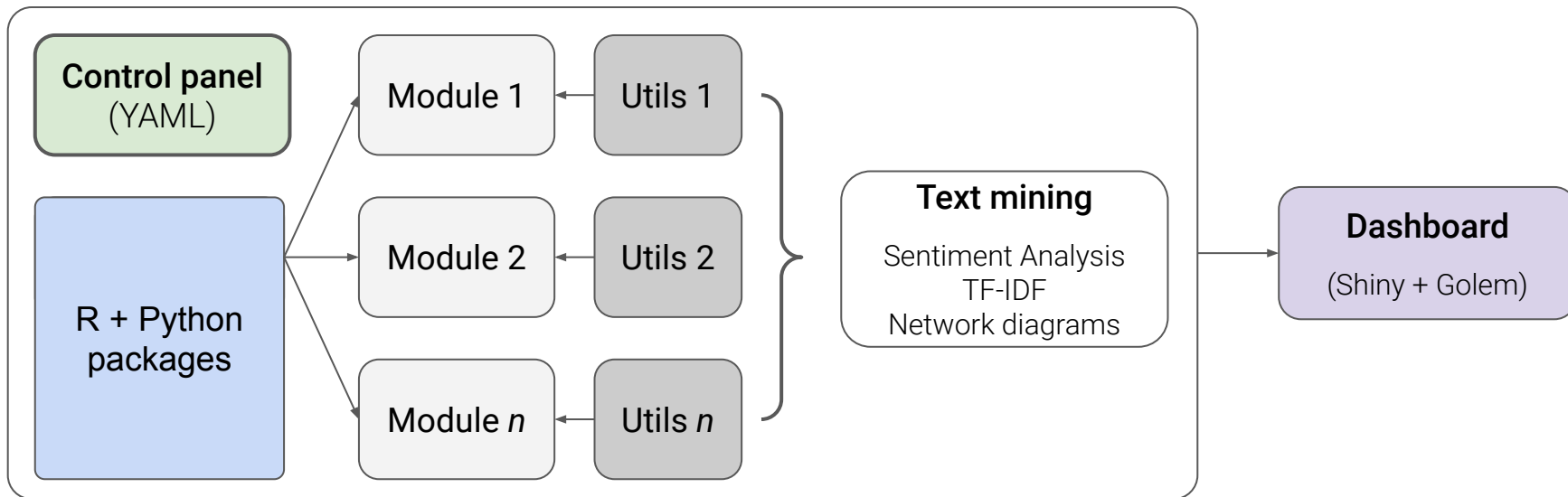
What is Golem?

“{golem} is an opinionated framework for building production-grade shiny applications.” (<https://github.com/ThinkR-open/golem>)

What problems can Golem solve?

1. We write complex apps that need to work similar with datasets from different NHS trusts.
2. We need a more formal framework to eliminate ambiguity and bugs in code.

How we use {golem}



From this...

```
68 - net_sentiment_nrc <- reactive({
69 -   text_data_filtered() %>%
70 -     tidytext::unnest_tokens(word, feedback) %>%
71 -     dplyr::left_join(tidytext::get_sentiments("nrc"), by = "word") %>% # We w
72 -     dplyr::count(linenumber, sentiment, name = "sentiment_count") %>%
73 -     dplyr::mutate(
74 -       sentiment_count =
75 +       dplyr::case_when(
76 -         is.na(sentiment) ~ NA_integer_,
77 -         TRUE ~ sentiment_count
78 -       )
79 -     ) %>%
80 -     dplyr::select(linenumber, sentiment, sentiment_count) %>%
81 -     tidyrr::pivot_wider(names_from = sentiment,
82 -       values_from = sentiment_count,
83 -       values_fill = 0,
84 -       names_sort = TRUE
85 -     ) %>%
86 -     dplyr::left_join(text_data_filtered(), by = "linenumber") %>%
87 -     dplyr::select(feedback, everything(), -"NA") %>%
88 -     # dplyr::mutate(all_sentiments =
89 -     #   dplyr::select(., dplyr::all_of(nrc_sentiments)) %>%
90 -     #   split(seq(nrow(.))) %>%
91 -     #   lapply(function(x) unlist(names(x)[x != 0]))
92 -     # ) %>%
93 -     #dplyr::select(feedback, all_sentiments, everything())
94 -     dplyr::select(feedback, everything())
95 - })
```

... to this

```
54 + net_sentiment_wide_nrc <- reactive({
55 +   experienceAnalysis::get_net_sentiment_wide_nrc(
56 +     text_data,
57 +     class_col_name = "label",
58 +     org_col_name = "organization",
59 +     filter_class = input$class,
60 +     filter_organization = input$organization)
66 61  })
```



lots of data tidying and analysis

**undocumented, untested, and
unavailable functions**

**R / Python pkgs + testing + open
source**

{nottshcPublic}



{experienceAnalysis}



Helper funs specific to
experiences data

{pxtextmining}



Text classification of NHS
patient feedback

{outcomeAnalysis}



Helper funs specific to
clinical outcomes data

{pxtextminingdashboard}



Text mining and visualisation
of NHS patient feedback

{experiencesdashboard}

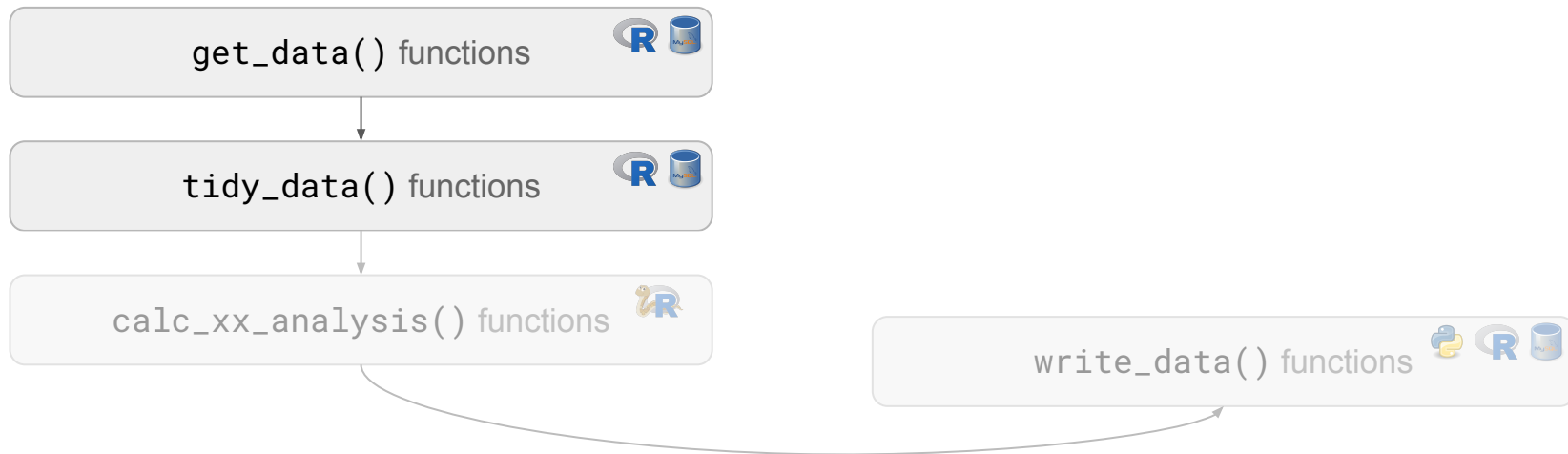
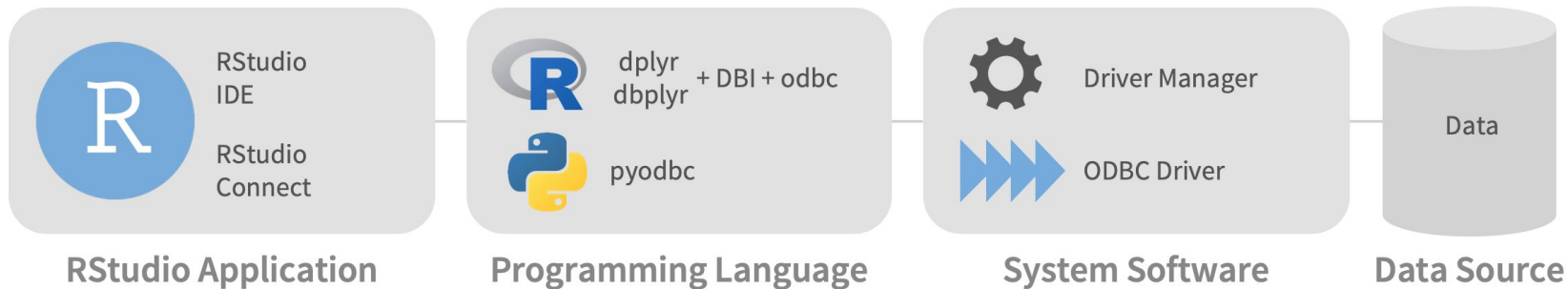


Helper funs specific to
experiences data

{honos}



Everything HoNOS

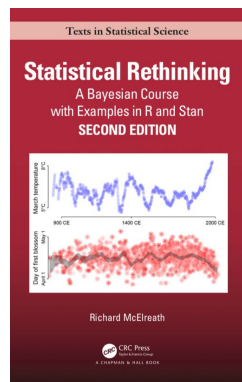


{nottshcPublic}

Demonstration

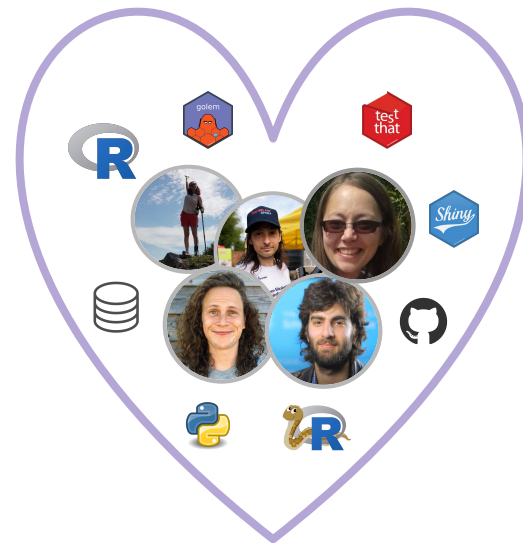


Book club



Code review

```
i <- 2  
  
while(i > 1) {  
  print(i)  
}
```



- ☐ **consider renaming this to nottshc + something else** [help wanted](#) [question](#)
#131 opened 4 days ago by milanwiedemann

Developing ideas

Developing as a team

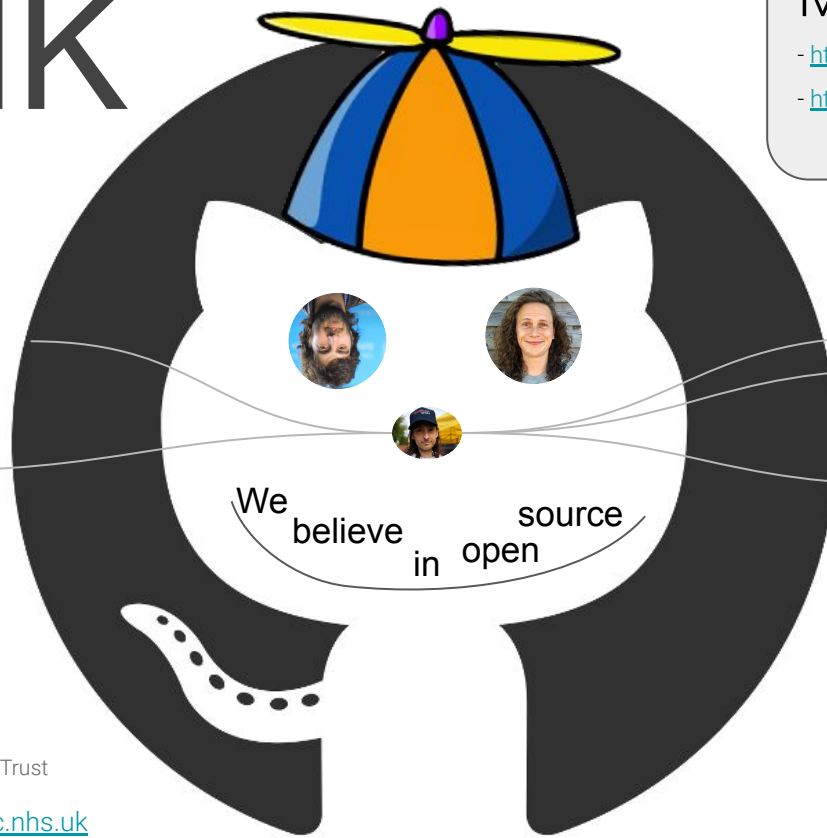
Developing tools

*“All technology problems
are people problems at
heart.”*

*“Open source is useful for
people to see, but it’s not
necessarily deployable.”*



THANK YOU



More about the team:

- <https://cdu-data-science-team.github.io/team-blog>
- <https://github.com/CDU-data-science-team>

Andreas Soteriades

Data Scientist 

Nottinghamshire Healthcare NHS Foundation Trust

Email: andreas.soteriades@nottshc.nhs.uk

GitHub: [andreassot10](https://github.com/andreassot10)

Milan Wiedemann

Data Scientist 

Nottinghamshire Healthcare NHS Foundation Trust

Email: milan.wiedemann@nottshc.nhs.uk

GitHub: [milanwiedemann](https://github.com/milanwiedemann)

Twitter: [@milanwiedemann](https://twitter.com/milanwiedemann)