MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

@kamalean_

Extracting Spatio-Temporal Trends in Medical Research Prioritization Through Natural Language Processing of Case Report Abstracts

Lean Franzl Yao, Kongmeng Liew Ph.D., Shoko Wakamiya Ph.D., Eiji Aramaki Ph.D.

Social Computing Laboratory

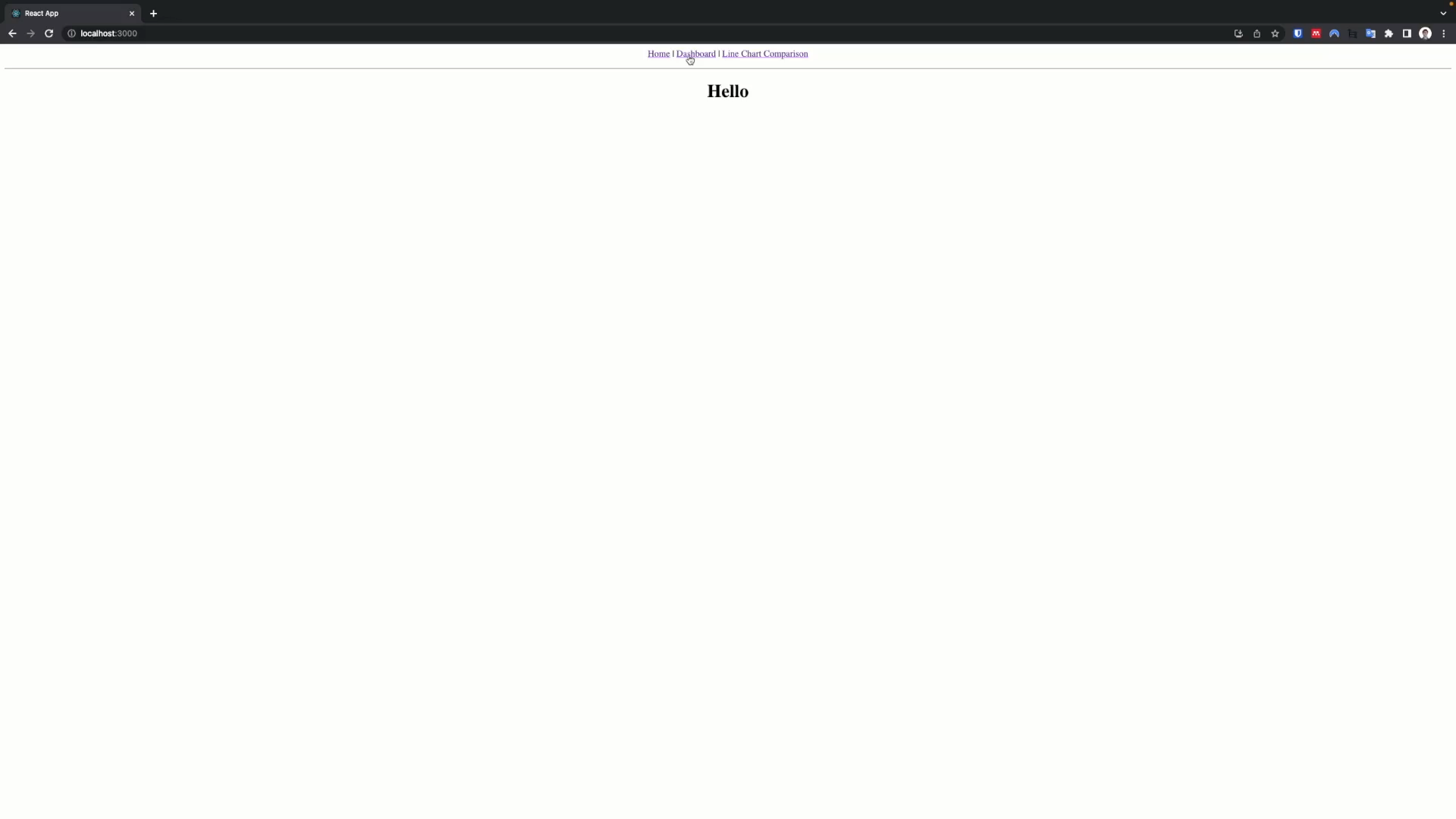
Nara Institute of Science & Technology

@TheInstituteDH #MEDINFO23

AIDH IMIA

Natural Language Processing (NLP) in Medicine

- Analyzing clinical documentation
  - Summarization, information extraction
- Coding medical terms
- Monitoring and detecting adverse drug events
- Meta-analysis
- Chatbots
- Pandemic and health measure surveillance


React App
localhost:3000
Home | Dashboard | Line Chart Comparison
Hello

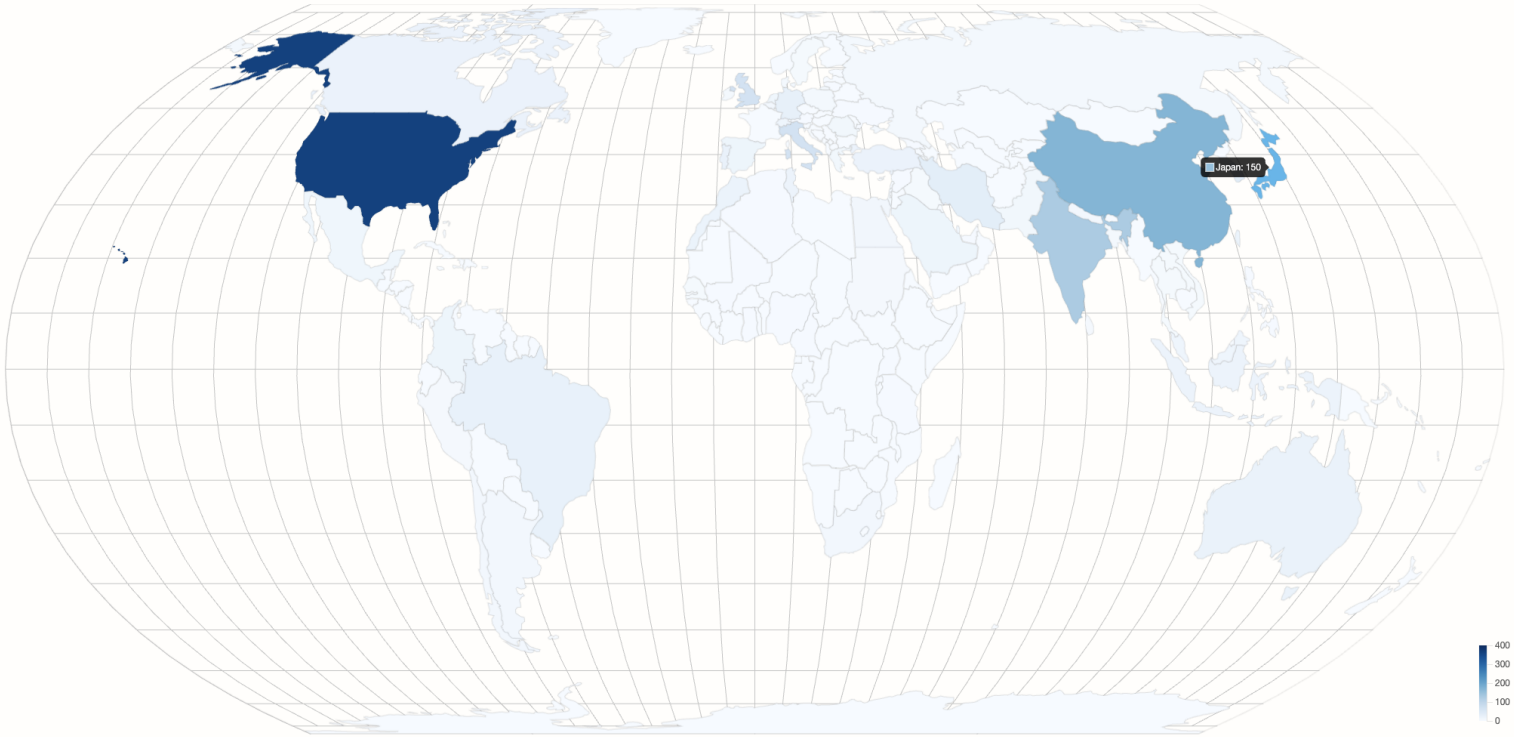
Example Use Case

- We want to do a study about infectious diseases
- What about Japan?
- How does Japan compare to Australia?

2021H1 | Japan | A00-B99: Certain infectious ar

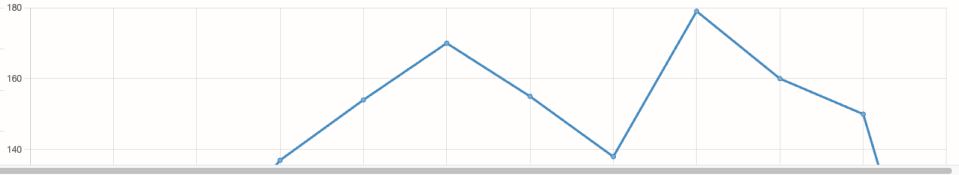
Change to Normalized Counts

Country distribution of A00-B99 case reports during 2021H1



ICD distribution of case reports in Japan during 2021H1

Trend of A00-B99 case reports in Japan



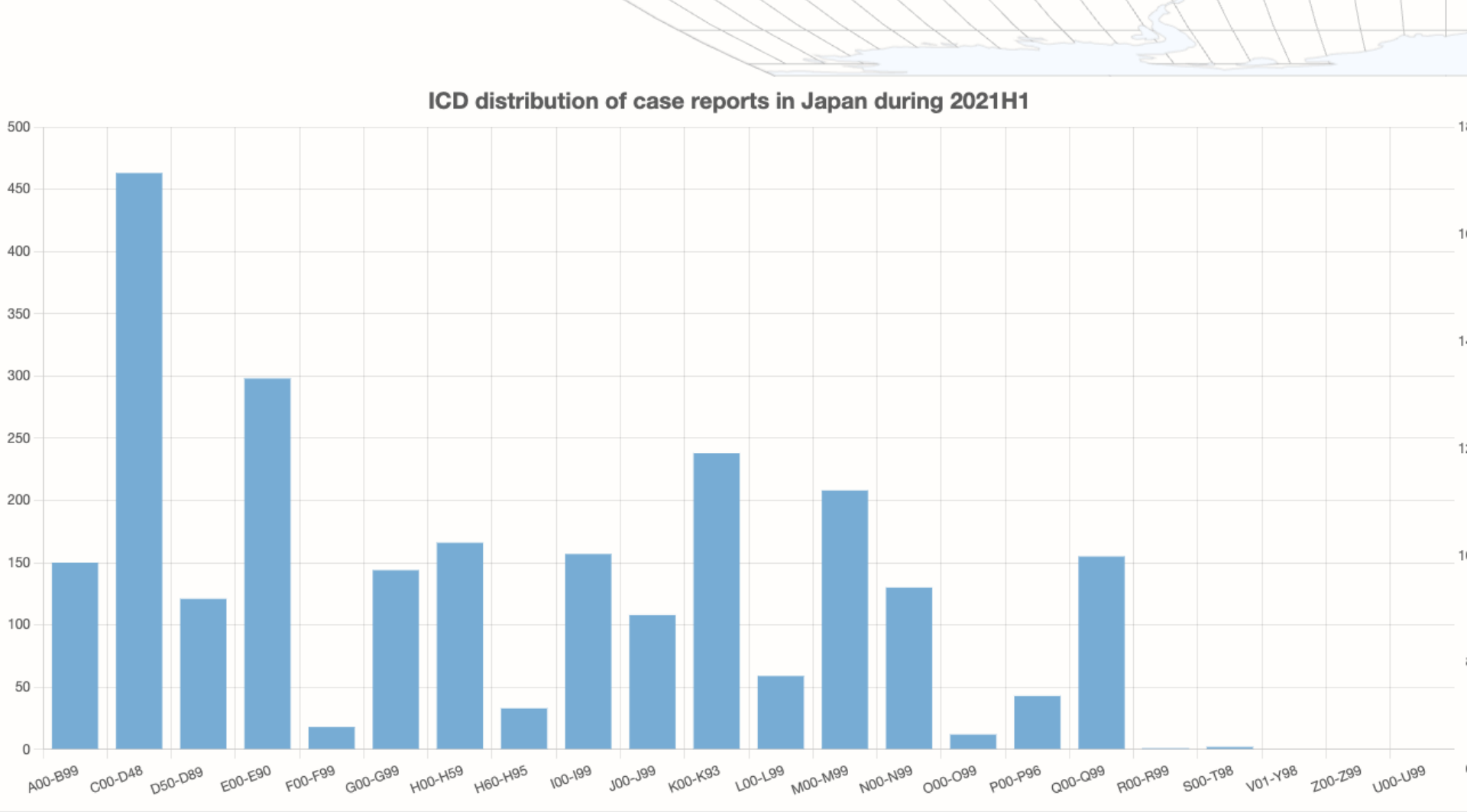
2021H1 | Japan | A00-B99: Certain infectious ar

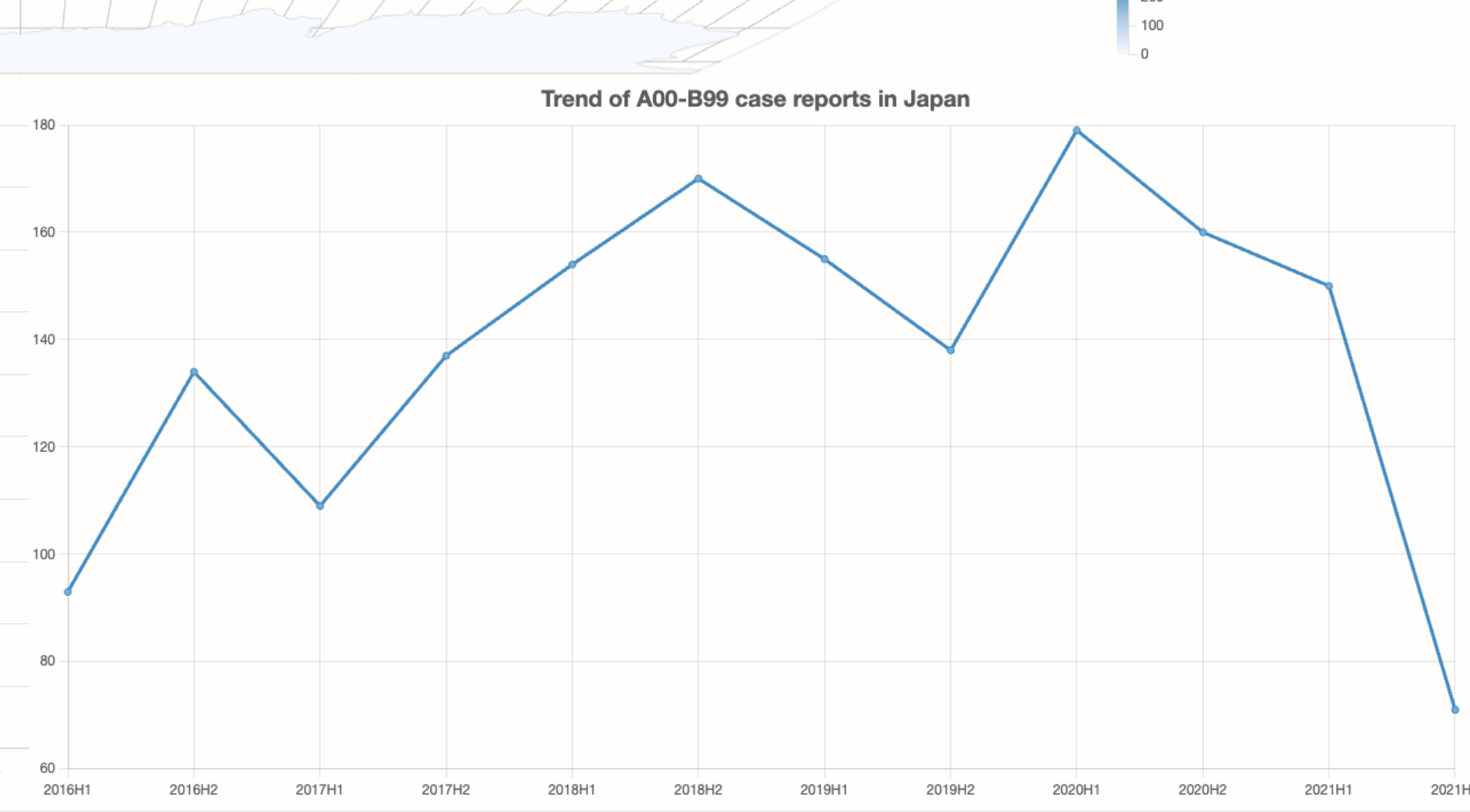
Change to Normalized Counts

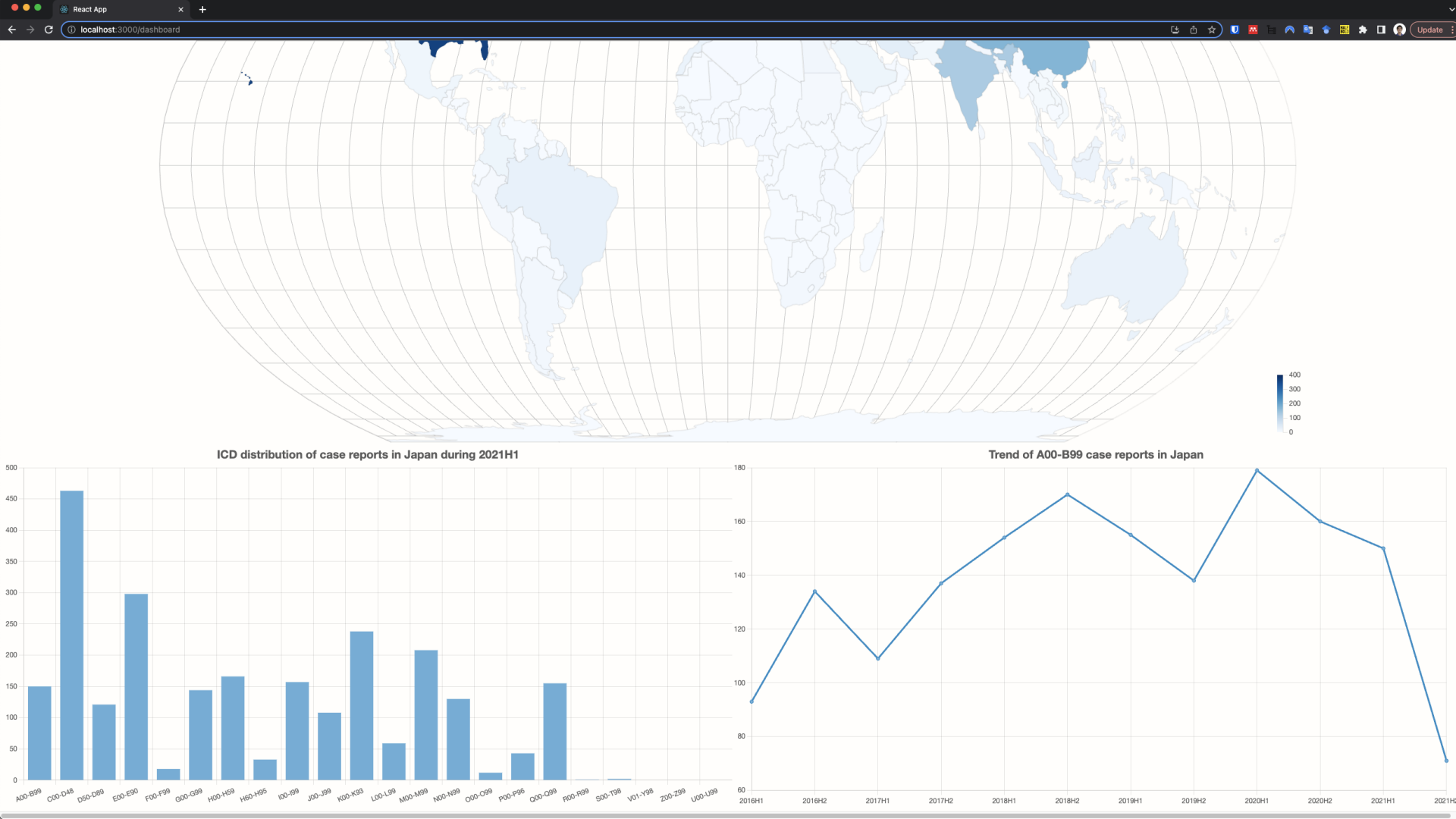
Country distribution of A00-B99 case reports during 2021H1

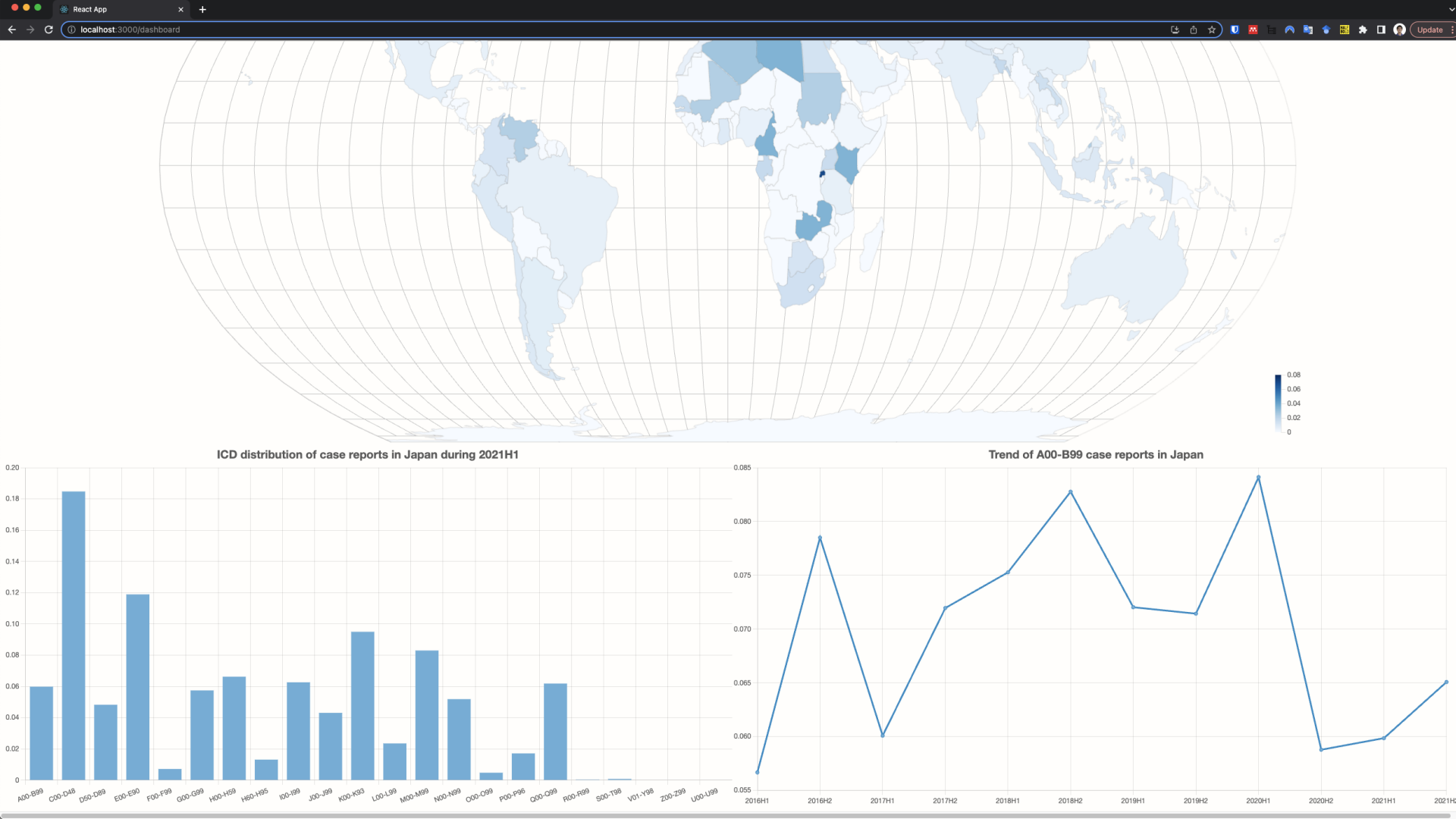
apan A00-B99: Certain infectious ar
Change to Normalized Counts

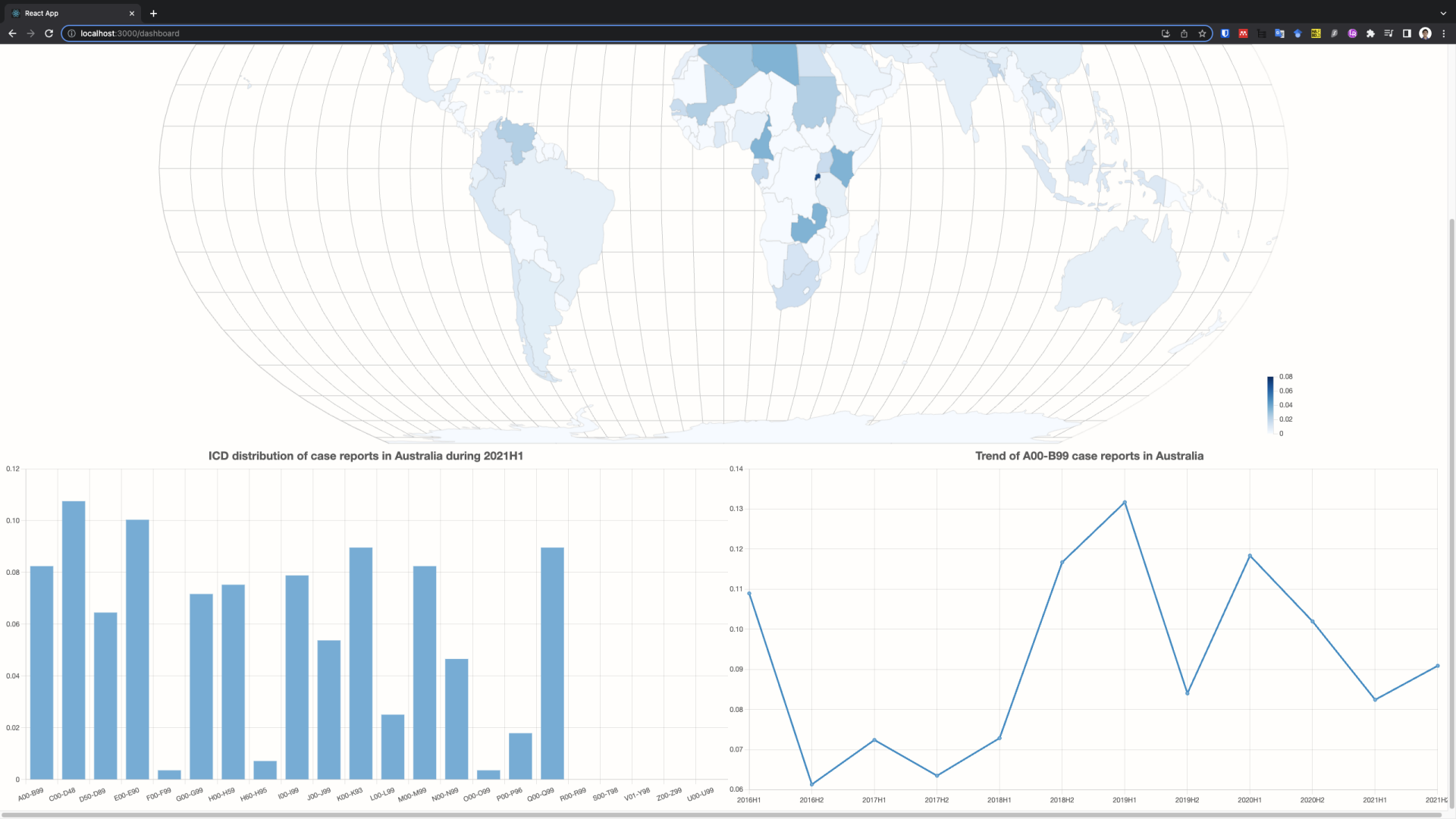
of A00-B99 case reports during 2021H1


ICD distribution of case reports in Japan during 2021H1
500
450
400
350
300
250
200
150
100
50
0
A00-B99
C00-D48
D50-D89
E00-E90
F00-F99
G00-G99
H00-H59
H60-H95
I00-I99
J00-J99
K00-K93
L00-L99
M00-M99
N00-N99
O00-O99
P00-P96
Q00-Q99
R00-R99
S00-T98
V01-Y98
Z00-Z99
U00-U99


100
0
Trend of A00-B99 case reports in Japan
180
160
140
120
100
80
60
2016H1
2016H2
2017H1
2017H2
2018H1
2018H2
2019H1
2019H2
2020H1
2020H2
2021H1
2021H

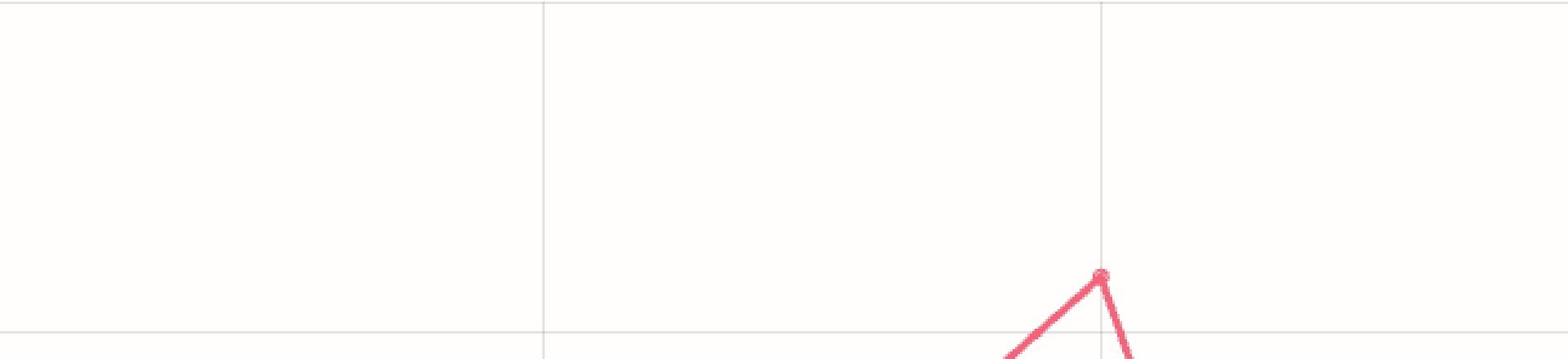

React App
localhost:3000/dashboard
Update
400
300
200
100
0
ICD distribution of case reports in Japan during 2021H1
500
450
400
350
300
250
200
150
100
50
0
A00-B99
C00-D48
D50-D89
E00-E90
F00-F99
G00-G99
H00-H59
H60-H95
I00-I99
J00-J99
K00-K93
L00-L99
M00-M99
N00-N99
O00-O99
P00-P96
Q00-Q99
R00-R99
S00-T98
V01-Y98
Z00-Z99
U00-U99
Trend of A00-B99 case reports in Japan
180
160
140
120
100
80
60
2016H1
2016H2
2017H1
2017H2
2018H1
2018H2
2019H1
2019H2
2020H1
2020H2
2021H1
2021H2

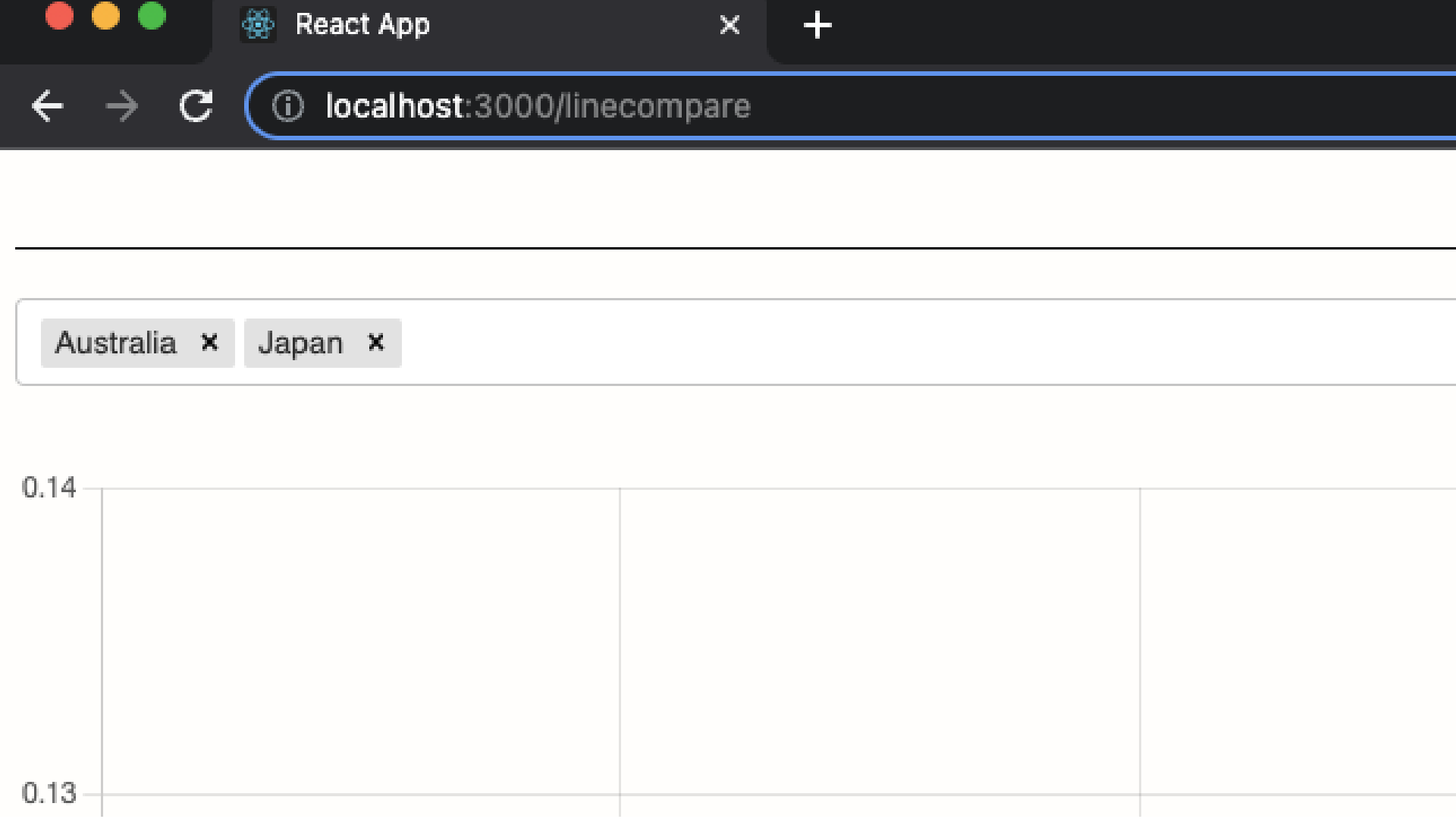

React App
localhost:3000/dashboard
Update
0.08
0.06
0.04
0.02
0
ICD distribution of case reports in Japan during 2021H1
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0
A00-B99
C00-D48
D50-D89
E00-E90
F00-F99
G00-G99
H00-H59
H60-H95
I00-I99
J00-J99
K00-K93
L00-L99
M00-M99
N00-N99
O00-O99
P00-P96
Q00-Q99
R00-R99
S00-T98
V01-Y98
Z00-Z99
U00-U99
Trend of A00-B99 case reports in Japan
0.085
0.080
0.075
0.070
0.065
0.060
0.055
2016H1
2016H2
2017H1
2017H2
2018H1
2018H2
2019H1
2019H2
2020H1
2020H2
2021H1
2021H2


React App
localhost:3000/dashboard
0.08
0.06
0.04
0.02
0
ICD distribution of case reports in Australia during 2021H1
0.12
0.10
0.08
0.06
0.04
0.02
0
A00-B99
C00-D48
D50-D89
E00-E90
F00-F99
G00-G99
H00-H59
H60-H95
I00-I99
J00-J99
K00-K93
L00-L99
M00-M99
N00-N99
O00-O99
P00-P96
Q00-Q99
R00-R99
S00-T98
V01-Y98
Z00-Z99
U00-U99
Trend of A00-B99 case reports in Australia
0.14
0.13
0.12
0.11
0.10
0.09
0.08
0.07
0.06
2016H1
2016H2
2017H1
2017H2
2018H1
2018H2
2019H1
2019H2
2020H1
2020H2
2021H1
2021H2

Change to Counts | A00-B99: Certain infectious ar

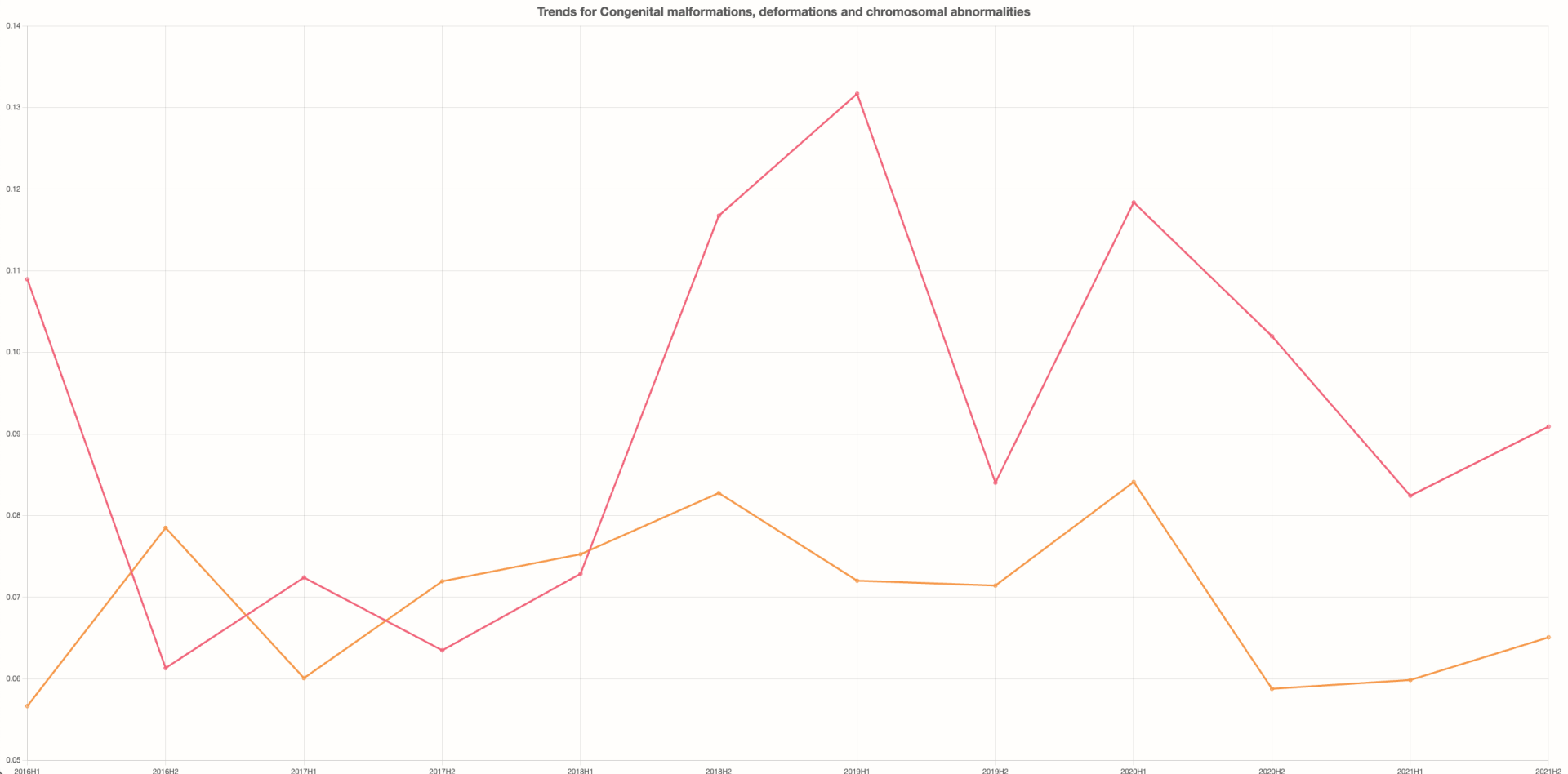
or Congenital malformations, deformations and chromosomal abn




React App
localhost:3000/linecompare
Australia
Japan
0.14
0.13

Change to Counts | A00-B99: Certain infectious ar

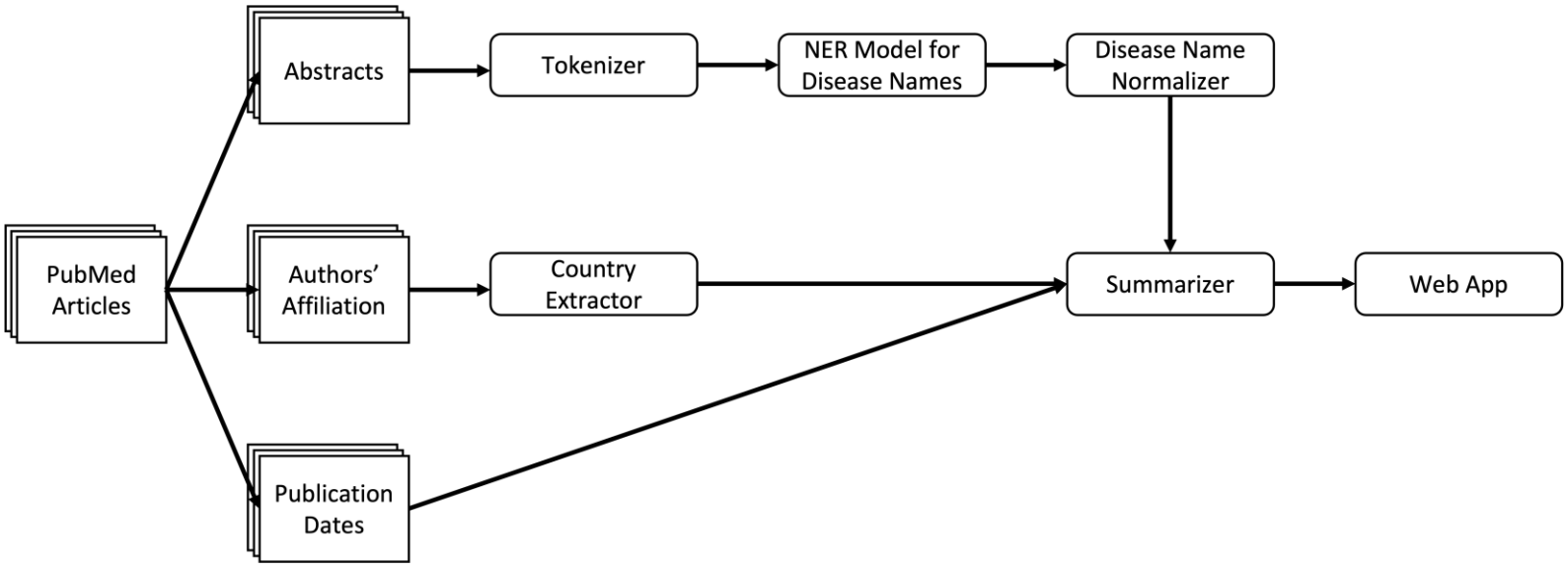
Australia × Japan ×



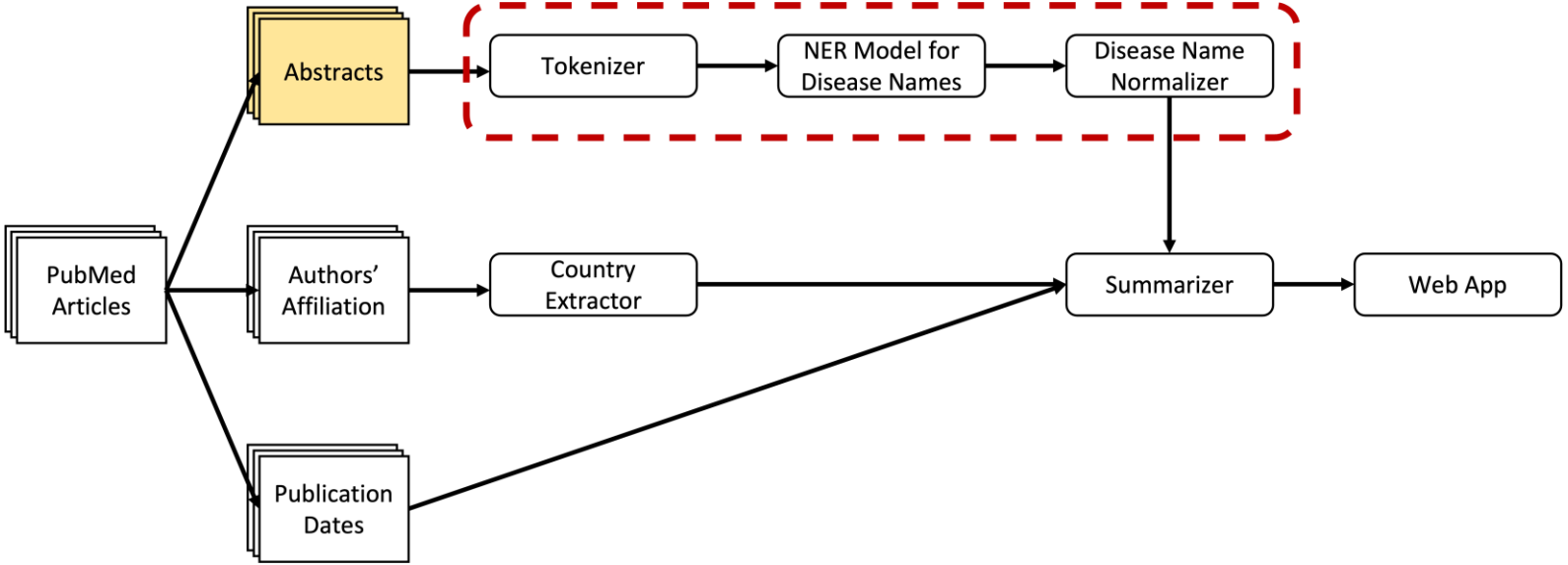
Data: PubMed Case Reports

- Contains important and novel information about
  - Symptoms
  - Diagnoses
  - Treatments
- 298,303 abstracts
- Search query: "case report"
- Date range: 2016-01-01 to 2021-10-31

System Architecture



Abstracts


Abstracts
Tokenizer
NER Model for Disease Names
Disease Name Normalizer
PubMed Articles
Authors' Affiliation
Country Extractor
Summarizer
Web App
Publication Dates

What is NER?

- Named-entity recognition (NER) is an information extraction task in NLP
- NER can be used to provide structure to text data
  - Extracting names of diseases, symptoms, treatments, etc.
  - Monitoring medications and dosages
  - Finding eligible patients for clinical trials

What is NER?

- Approaches to NER
  - Deep learning models
  - Non-neural network machine learning models ← our approach
- Treated as a sequence-labeling problem

I feel sick, maybe I have dengue fever.

I	feel	sick	maybe	I	have	dengue	fever
O	O	O	O	O	O	Disease	Disease

Results

Human Cytomegalovirus (HCMV) still represents a crucial concern in solid organ transplant recipients (SOTRs) and the use of antiviral therapy are limited by side effects and the selection of viral mutations conferring antiviral drug resistance.Here we reported the case of an HCMV seronegative patient with common variable immunodeficiency (CVID), multiple hepatic adenomatosis, hepatopulmonary syndrome and portal hypertension who received a liver transplant from an HCMV seropositive donor. The patient was treated with Valganciclovir (vGCV) and then IV Ganciclovir (GCV) at 5 week post-transplant for uncontrolled HCMV DNAemia. However, since mutation A594V in UL97 gene conferring resistance to ganciclovir was reported, GCV therapy was interrupted. Due to the high toxicity of Foscarnet (FOS) and Cidofovir (CDV), Letermovir (LMV) monotherapy at the dosage of 4800mg per day was administered, with a gradual viral load reduction. However, a relapse of HCMV DNAemia revealed the presence of mutation C325Y in HCMV UL56 gene conferring resistance to LMV.In conclusion, even if LMV is an effective and favorable safety molecule it might have a lower genetic barrier to resistance. A warning on the use of LMV monotherapy as rescue treatments for HCMV GCV-resistant infections in transplant recipients is warranted.© 2021. The Author(s).

Results

Emergence of Letermovir-resistant HCMV UL56 mutant during rescue treatment in a liver transplant recipient with ganciclovir-resistant infection HCMV: a case report

PMID: 34556034

ICD: K76.81

Hepatopulmonary syndrome

Human Cytomegalovirus (HCMV) still represents a crucial concern in solid organ transplant recipients (SOTRs) and the use of antiviral therapy are limited by side effects and the selection of viral mutations conferring antiviral drug resistance.Here we reported the case of an HCMV seronegative patient with common variable immunodeficiency (CVID), multiple hepatic adenomatosis, hepatopulmonary syndrome and portal hypertension who received a liver transplant from an HCMV seropositive donor. The patient was treated with Valganciclovir (vGCV) and then IV Ganciclovir (GCV) at 5 week post-transplant for uncontrolled HCMV DNAemia. However, since mutation A594V in UL97 gene conferring resistance to ganciclovir was reported, GCV therapy was interrupted. Due to the high toxicity of Foscarnet (FOS) and Cidofovir (CDV), Letermovir (LMV) monotherapy at the dosage of 4800mg per day was administered, with a gradual viral load reduction. However, a relapse of HCMV DNAemia revealed the presence of mutation C325Y in HCMV UL56 gene conferring resistance to LMV.In conclusion, even if LMV is an effective and favorable safety molecule it might have a lower genetic barrier to resistance. A warning on the use of LMV monotherapy as rescue treatments for HCMV GCV-resistant infections in transplant recipients is warranted.© 2021. The Author(s).

Results

- Performance on the NCBI Disease corpus
  - Precision: 0.8222
  - Recall: 0.7172
  - F1-Score: 0.7661
- Found mentions of disease names in 264,399 out of 298,303 abstracts (88.63%)
- 205,939 abstracts (69.04%) contained mention of disease names mapped to ICD-10-CM codes

Discussion and Limitations

- Classifying case reports according to disease categories, date, and country to monitor trends in medical research prioritization
- Can serve as a tool to help in research funding allocation and collaborative research
- Resulting data can be used with or as external data in other models and studies
- This preliminary study is still being continued

Discussion and Limitations

- Components in the system can be modified to fit the scope of interest or the relevant research problem
- State-of-the-art technologies could improve the results or efficiency of the system

Conclusions

- Used NLP tools to collect, classify, and summarized publicly available medical case reports according to date, country, and disease category
- Output can be used to make informed decisions on funding allocation and collaboration efforts
- Non-state-of-the-art approaches are still capable of producing good and usable insights