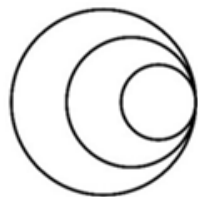


# How to design and deliver pathogen genomics training for health and research professionals

Workshop on Data Visualisation  
Telling a story and making a point  
Monica Abrudan, PhD  
7/3/2023



**wellcome  
connecting  
science**



Centre for Genomic  
Pathogen Surveillance



# Learning outcomes

- List three principles of good data viz
- Critique different alternative data analyses methods for producing data visualisations.
- Demonstrate how to use Microreact to tell a story

# Data Visualisation - Telling a story and making a point

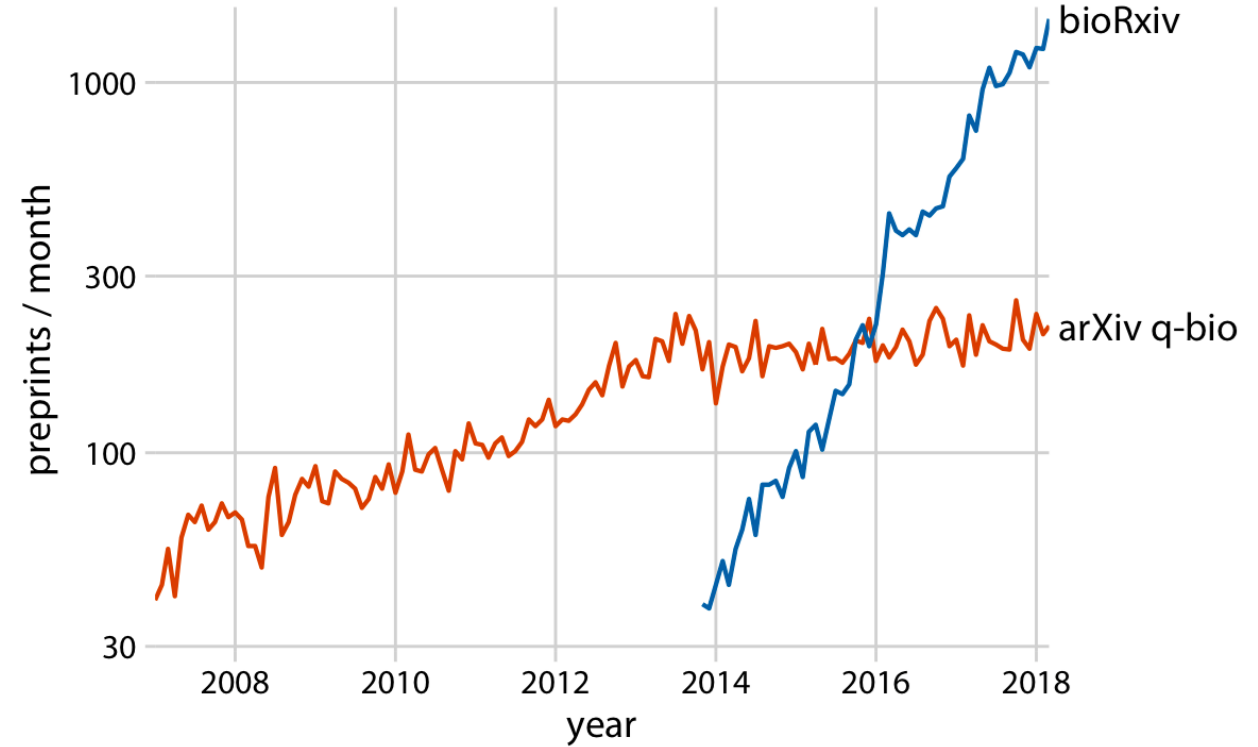
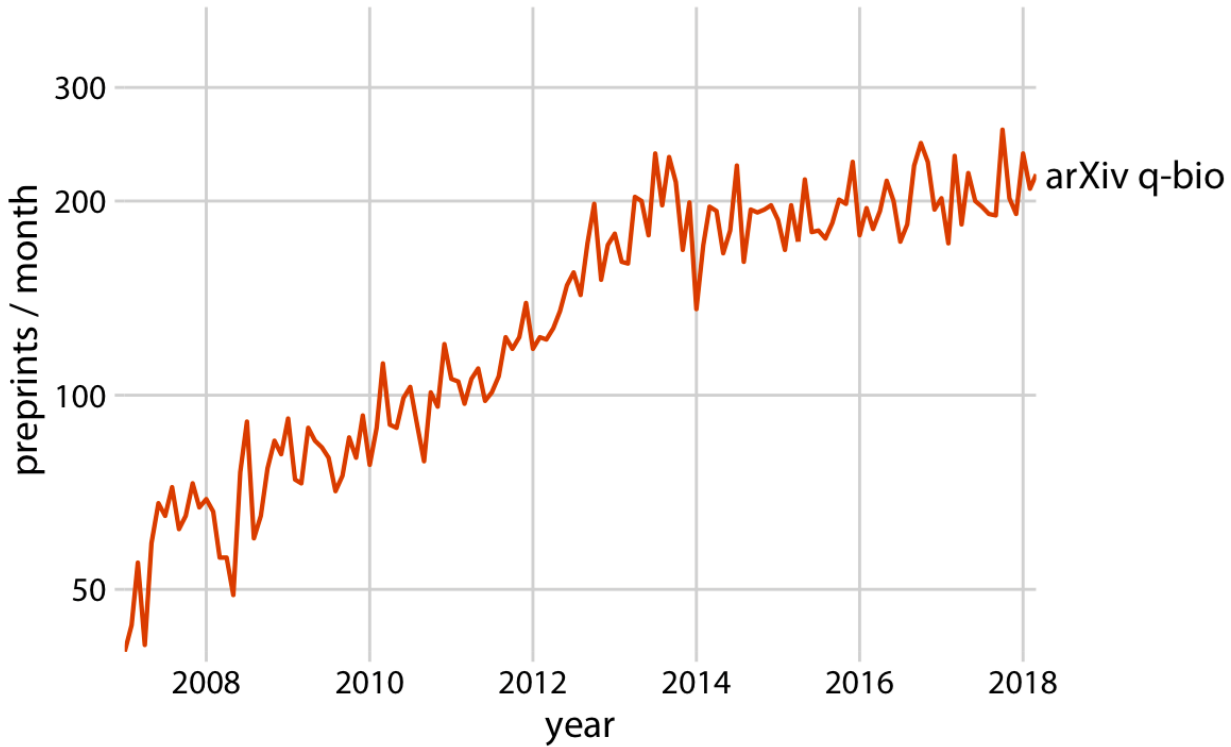
Telling a story with data

What is a story?

Opening - Challenge- Action - Resolution

# How we tell a story is important

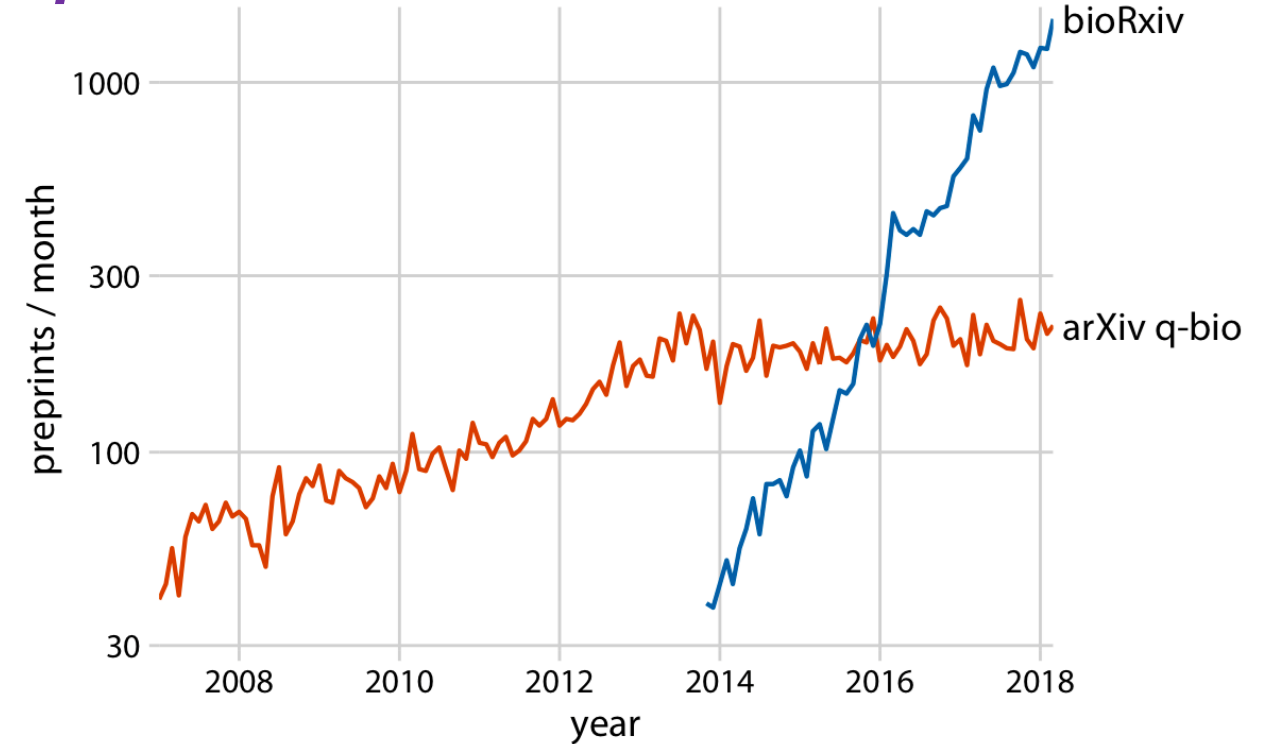
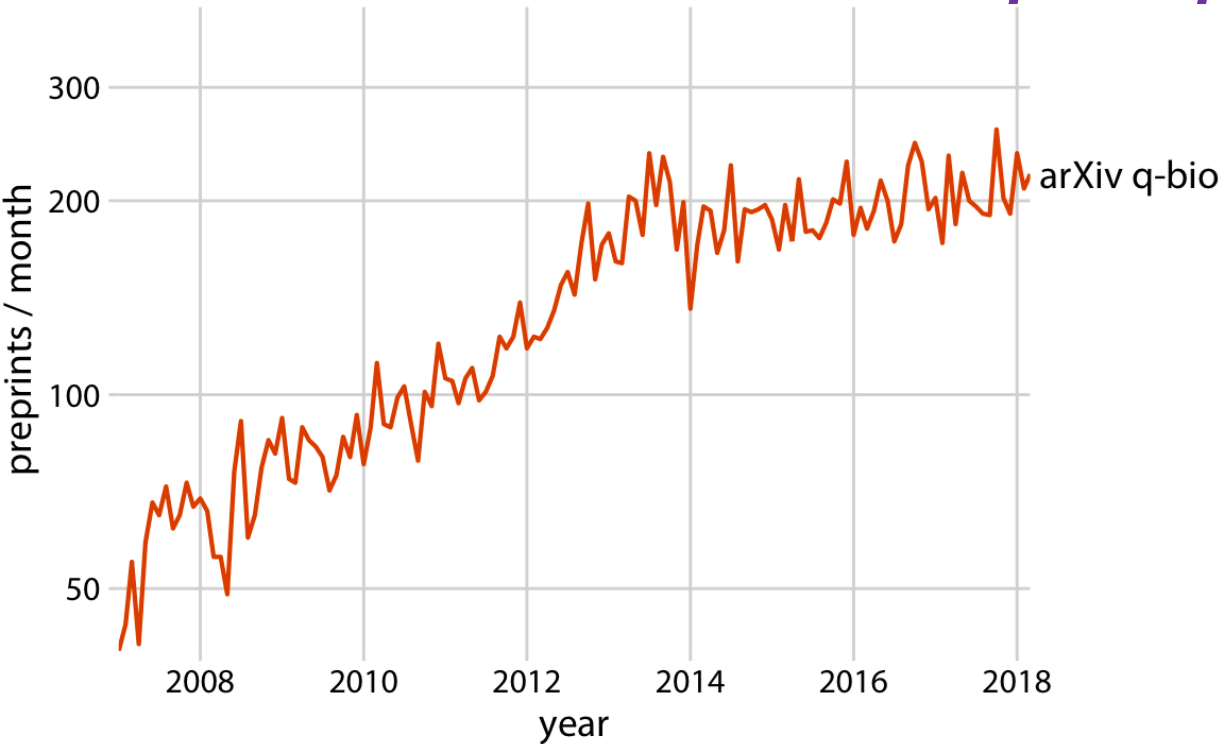
## *Example of pre-prints in science*



<https://clauswilke.com/dataviz/>

# How we tell a story is important

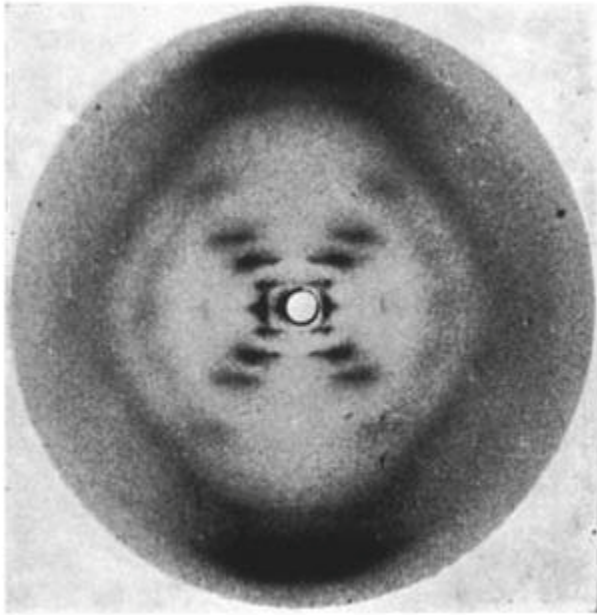
## *Example of pre-prints in science*



**The story has the strongest impact when broken into two pieces**

<https://clauswilke.com/dataviz/>

# Science Stories



Data



Information

**A Structure for Deoxyribose Nucleic Acid**  
**J. D. Watson and F. H. C. Crick (1)**

**April 25, 1953 (2), *Nature* (3), 171, 737-738**

Knowledge

# This Workshop

Data                      Information                      Knowledge



***Exercise:***

How to  
generate a plot  
from raw data

***Exercise:***

How to tell a  
story using  
Microreact

**Data**  **Information**

## Real scenario from a public health lab in the USA

*Question: Can we identify ways to reduce poor sequencing outcomes?*

A state public health lab receives samples from individual labs, in the format shown below:

Sample	Lab name	Source	Target	Collection date	Received date
Sample1	Alpha			02/02/2023	07/02/2023
Sample2	Bravo			04/02/2023	08/02/2023
Sample3	Charlie			06/02/2023	09/02/2023
Sample4	Alpha			01/02/2023	10/02/2023
Sample5	Bravo			08/02/2023	11/02/2023
Sample6	Charlie			07/02/2023	08/02/2023
Sample7	Alpha			31/01/2023	09/02/2023



Data  Information

## Real scenario from a public health lab in the USA

*Question: Can we identify ways to reduce poor sequencing outcomes?*

The state public health lab sequences and analyses the data. The results of the bioinformatics analyses look like this:

Sample	Date processed	Outcome
Sample1	14/02/2023	Pass
Sample2	15/02/2023	Fail
Sample3	16/02/2023	Pass
Sample4	17/02/2023	Pass
Sample5	18/02/2023	Fail
Sample6	14/02/2023	Pass
Sample7	15/02/2023	Pass
Sample8	16/02/2023	Fail

Data  Information

## Real scenario from a public health lab in the USA

*Question: Can we identify ways to reduce poor sequencing outcomes?*

We want to produce some figures that would allow us to 1) determine which labs are the slowest ones to submit data and 2) determine which labs have the poorest outcomes.



Data  Information

## Real scenario from a public health lab in the USA

*Question: Can we identify ways to reduce poor sequencing outcomes?*

- Three approaches to solve the problem:
  - [Excel data manipulation](#)
  - [Python in google colab](#)
  - [Data-flo](#) and [Microreact](#)
- Group discussions:
  - Which approach would you train others to apply and why ?
  - What happens if the data input changes?
  - What if we add more input data?
  - What if dates intervals change? (what if you have to select a subset of your initial data)
  - What if the question changes, would you change the methods to would teach you audience?
  - Can you suggest a different approach?

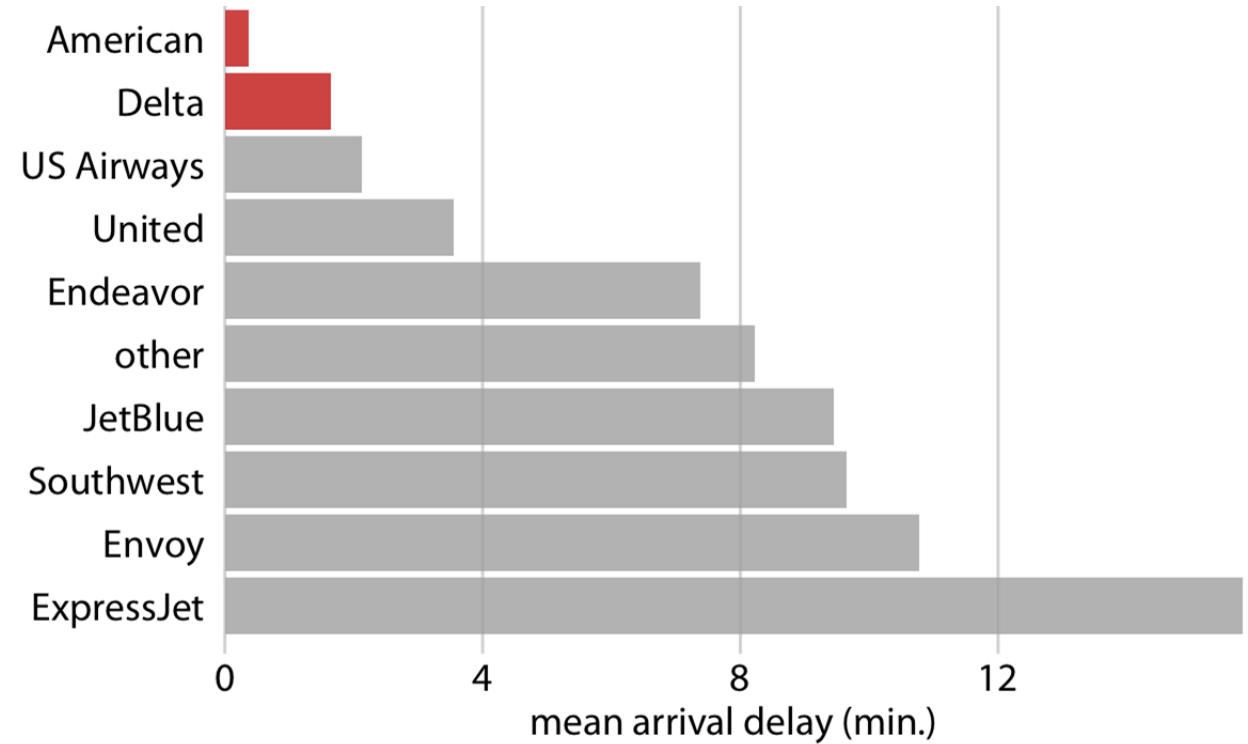
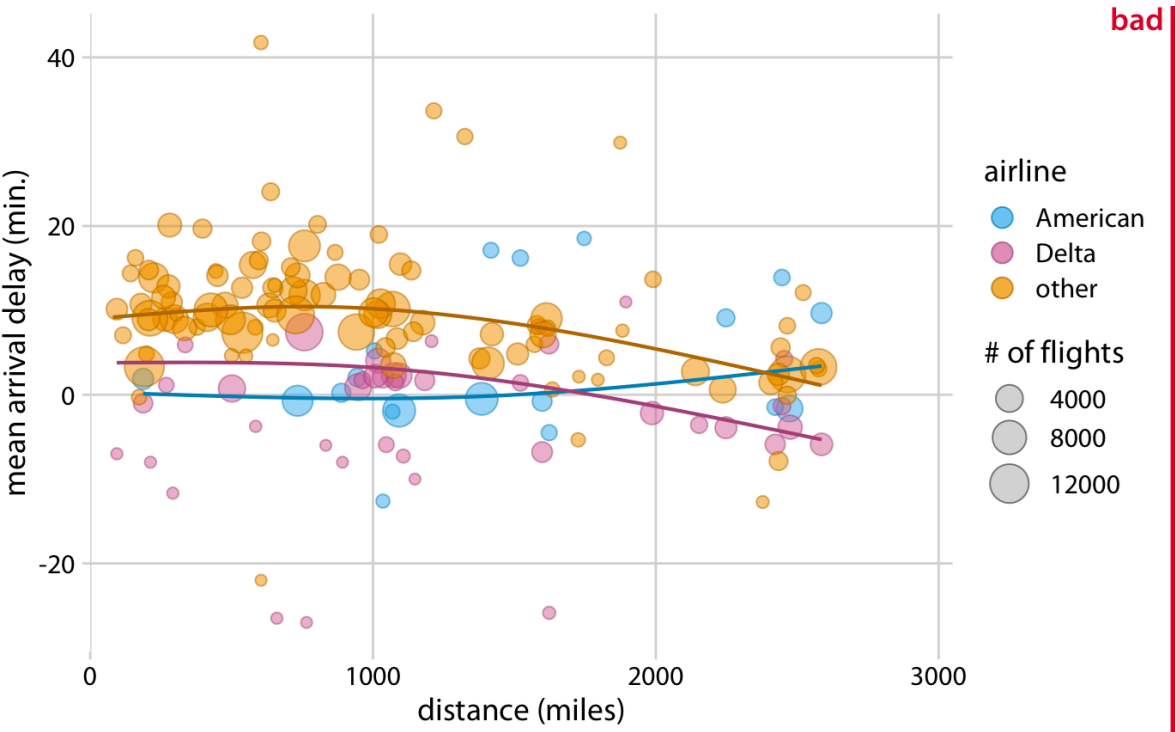
Information → Knowledge

## An example of using Microreact to tell a story

- A short Microreact demo, explaining basic features
- [An example of story telling using Microreact](#)
- Exercise: [Tell a story using Microreact](#)

# Final thoughts

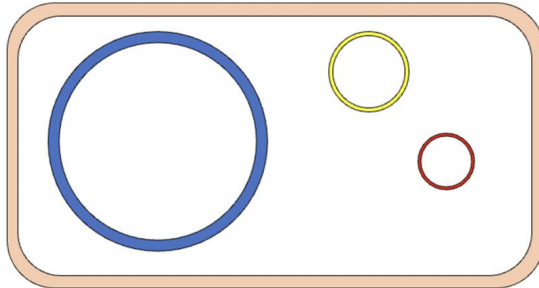
# Keep it simple, where possible



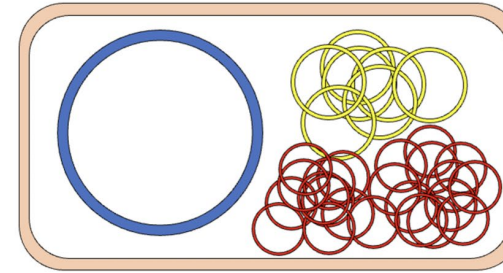
<https://clauswilke.com/dataviz/>

# Build up towards complex figures

Choice of methods

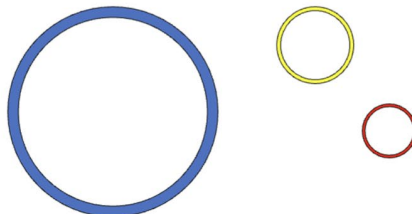
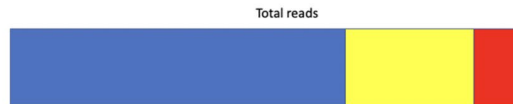


Choice of methods

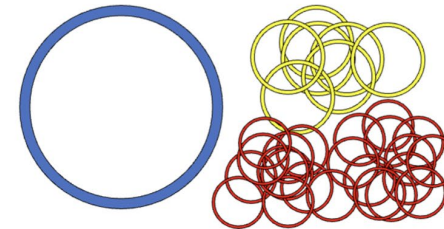


Dr Matthew Dorman

Choice of methods



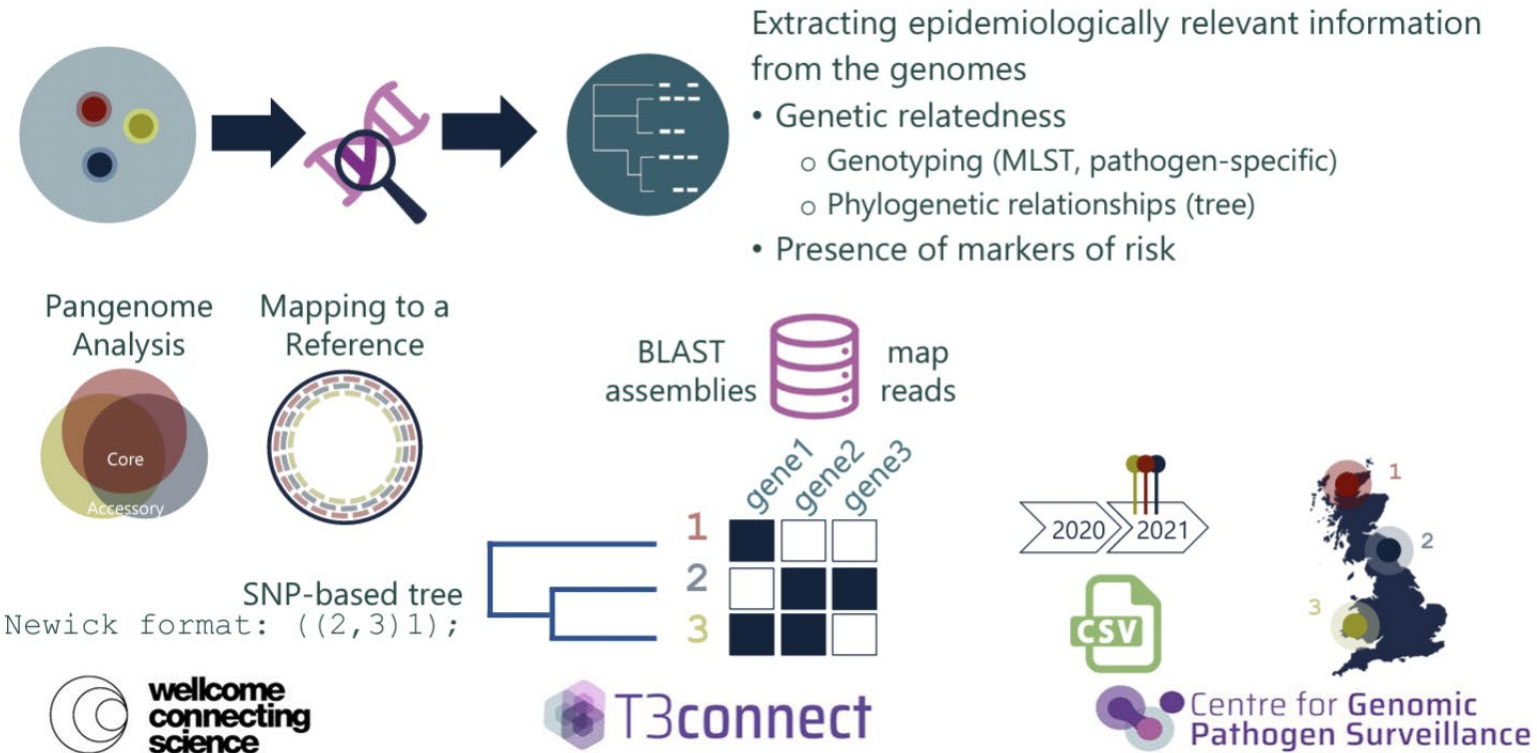
Choice of methods





# Build up towards complex figures

## Genomic Epidemiology in a Nutshell

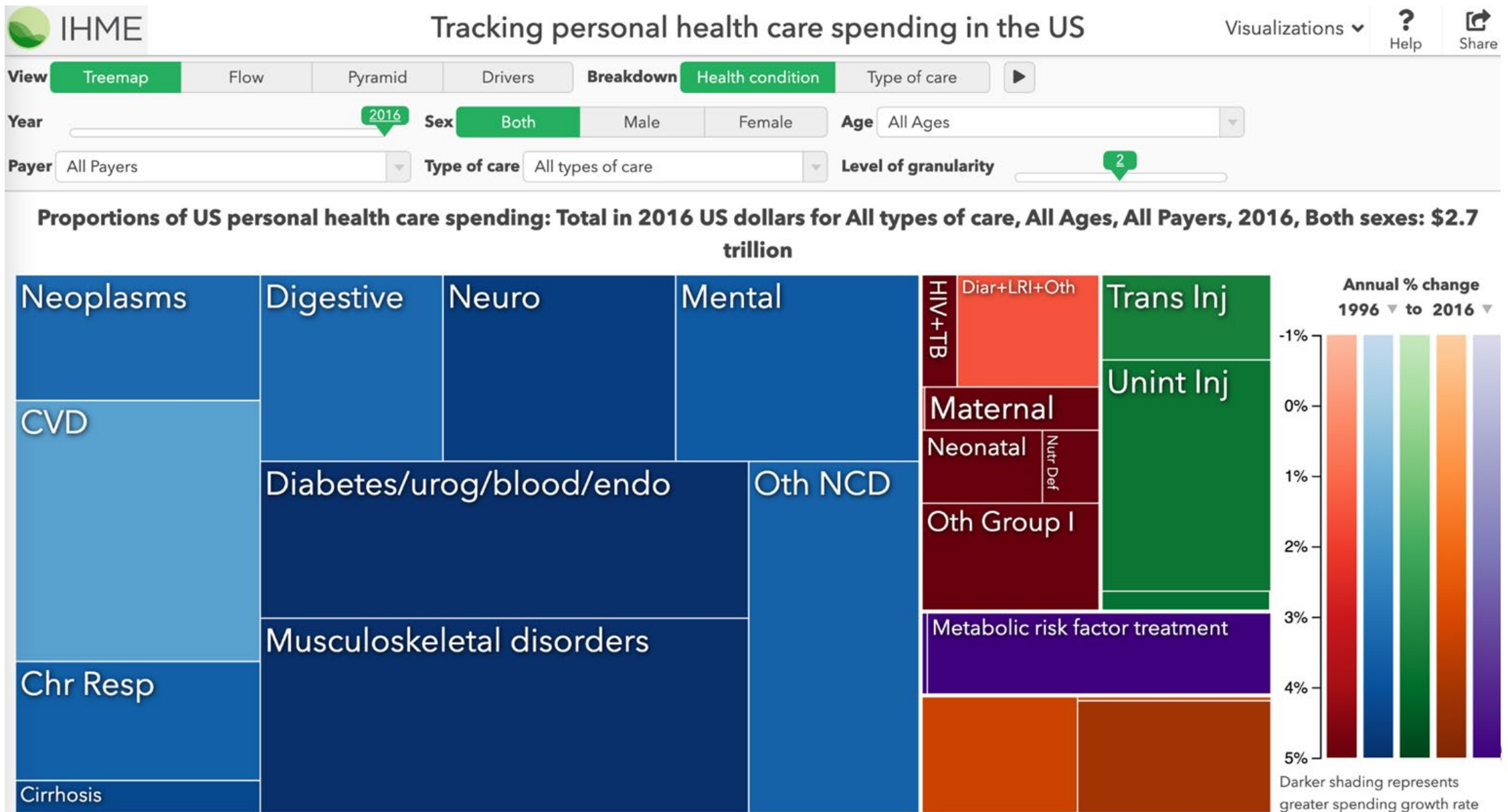


Dr Silvia Argimon



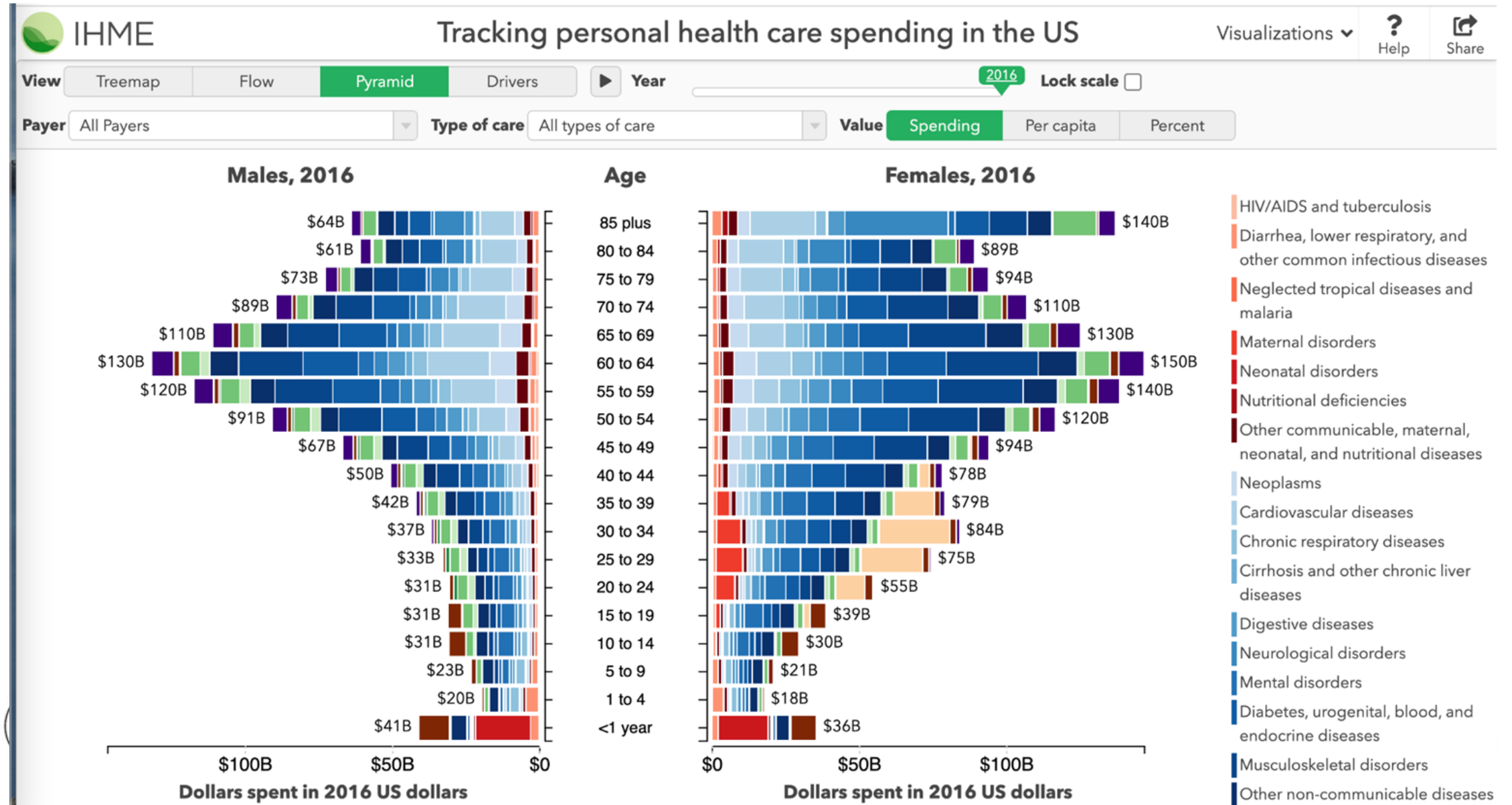
# Be consistent, but do not be repetitive

<https://vizhub.healthdata.org/dex/>



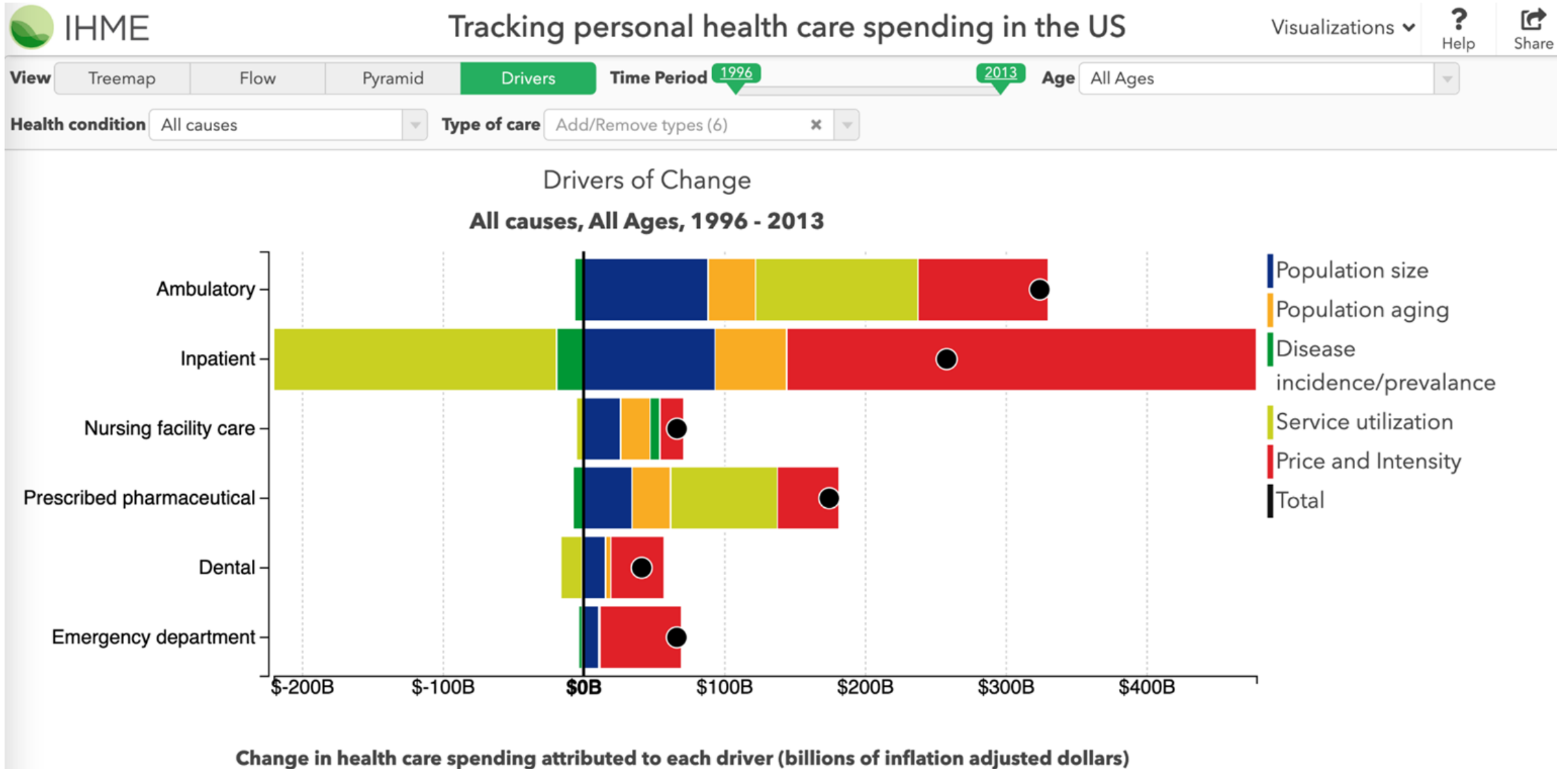
# Be consistent, but do not be repetitive

<https://vizhub.healthdata.org/dex/>



# Be consistent, but do not be repetitive

<https://vizhub.healthdata.org/dex/>

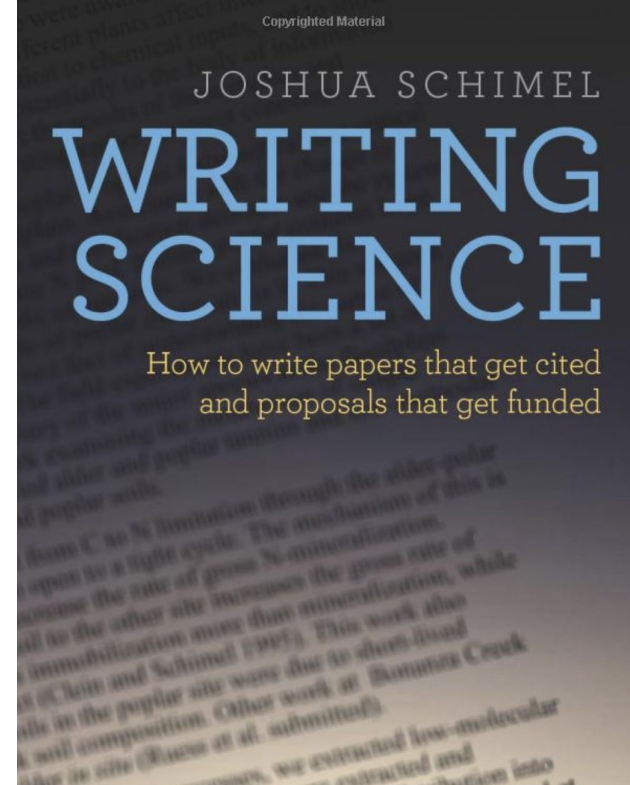




# Resources



- <https://ourworldindata.org/eradication-of-diseases>
- <https://vizhub.healthdata.org/microbe>
- <https://www.who.int/data/gho>
- <https://www.visualisingdata.com/>
- <https://clauswilke.com/dataviz/>
- <https://observablehq.com/>
- <https://www.worldpop.org/>
- <https://www.shipmap.org/>
- <https://www.healthdata.org/data-tools-practices/interactive-data-visuals>



# Acknowledgements

This course was developed by a collaboration between the [Centre for Genomic Pathogen Surveillance](#) and [Wellcome Connecting Science](#). It was brought to you by [T3Connect – Data Science and Genomic Pathogen Surveillance Training Programme](#), funded by [UKRI](#).

This module contains materials from the following sources:

- [Storyset | Customize, animate and download illustration for free](#)



# Creative Commons

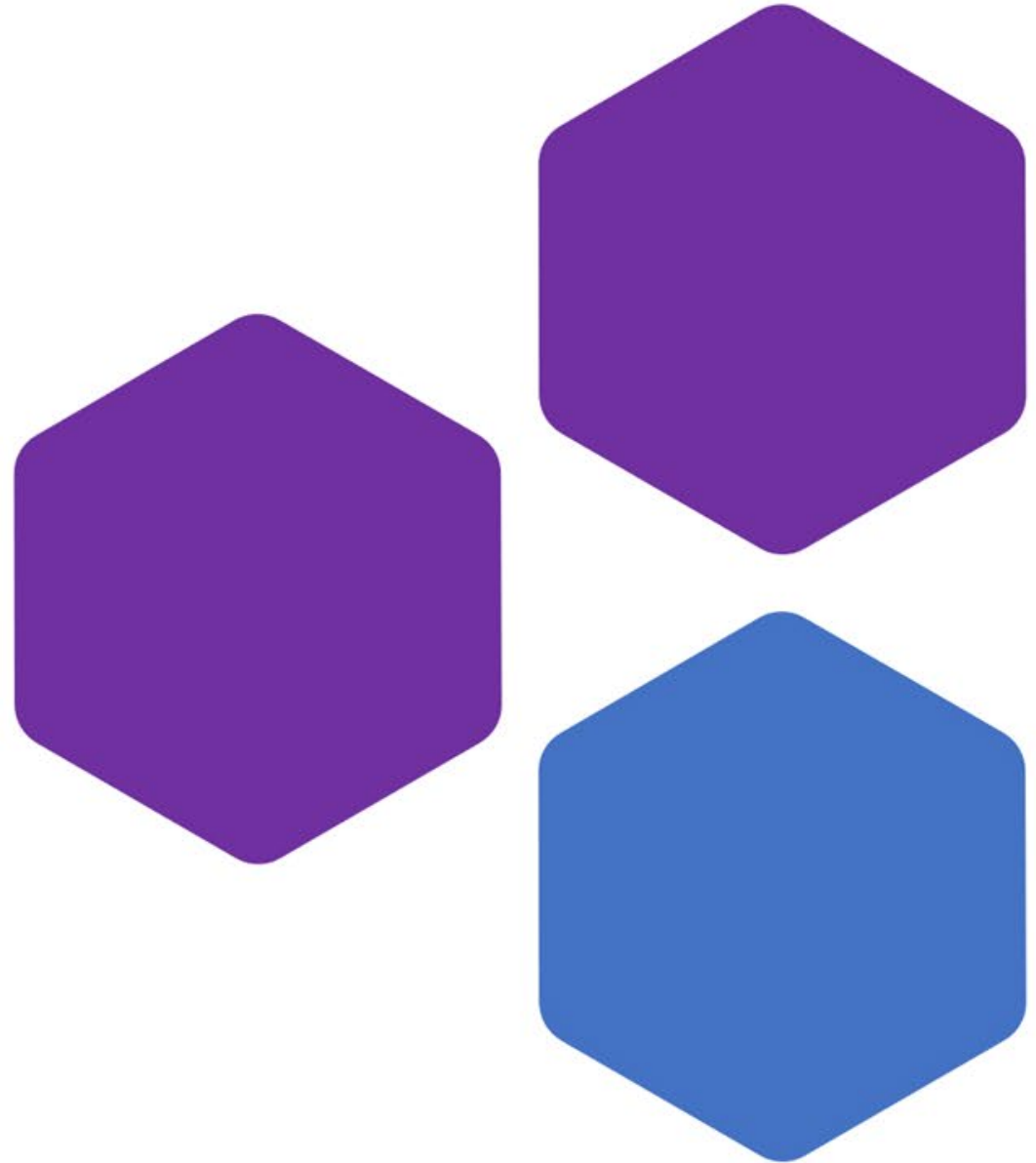
This work is licensed under a [Creative Commons Attribution-Share Alike Licence \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/).



**Attribution-ShareAlike 4.0 International  
(CC BY-SA 4.0)**

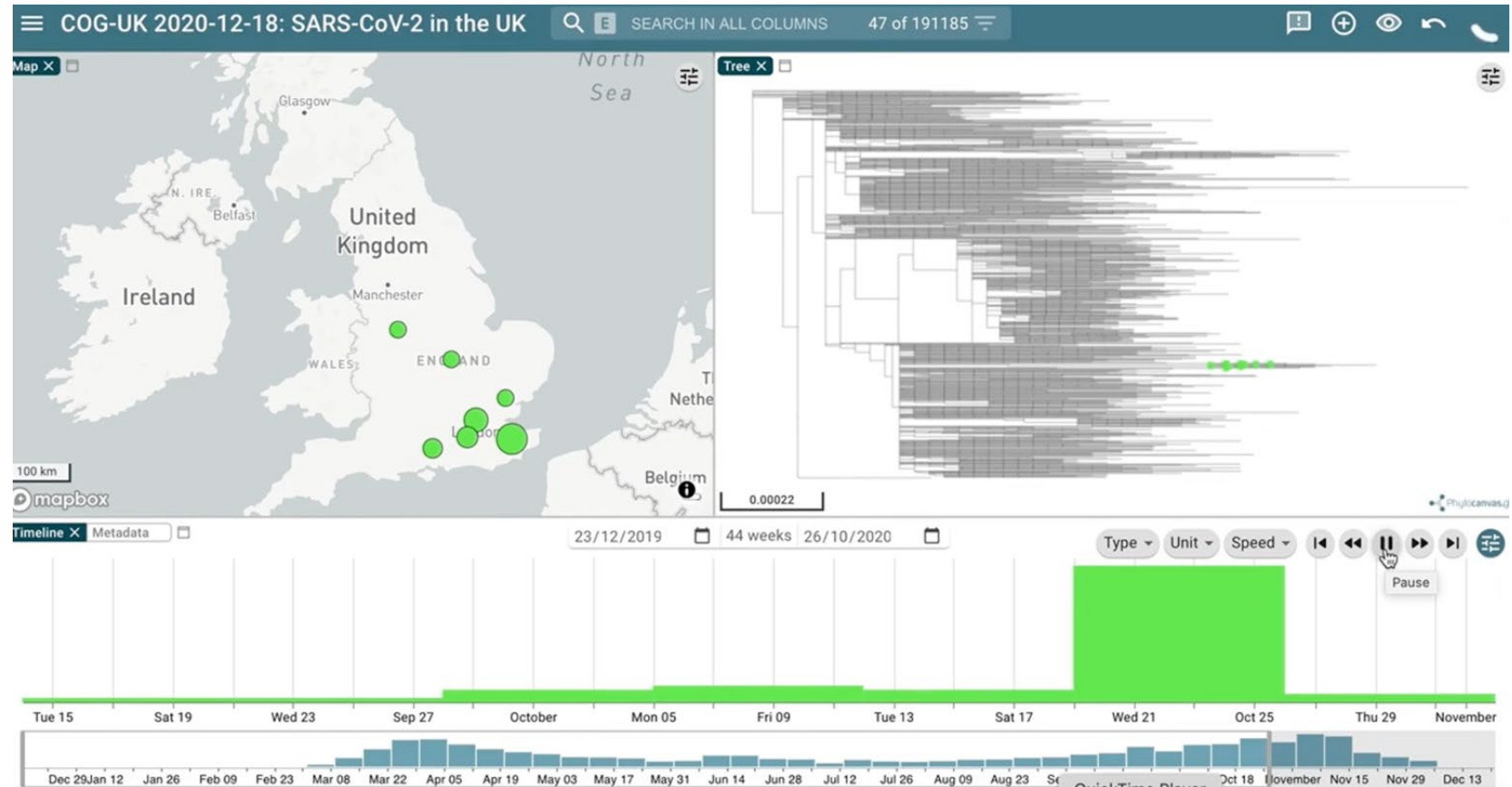


**Thank you**



# Re-designing a visualisation tool

Covid19 dashboards

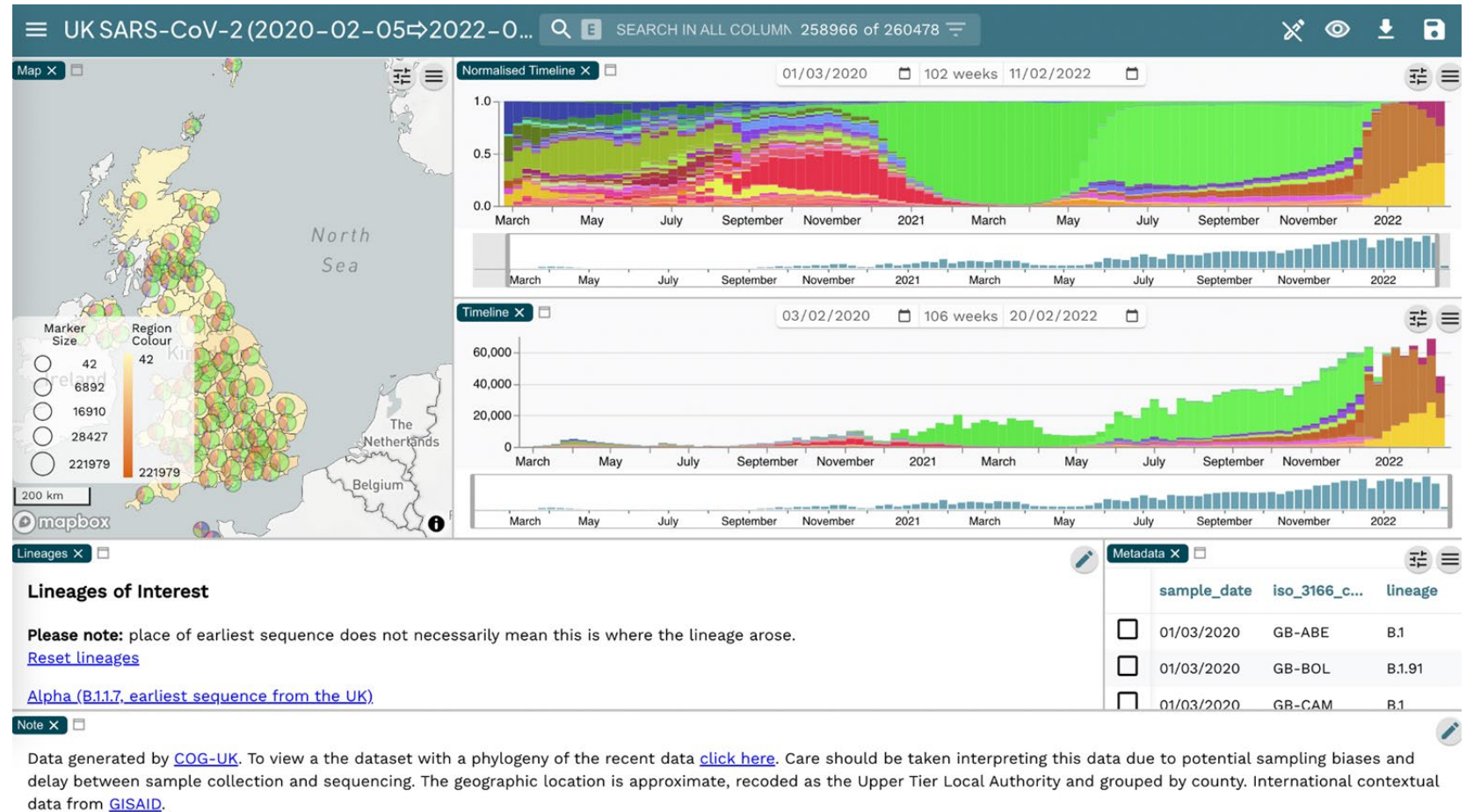


Microreact showing the emergence of SARS-CoV-2 Lineage B.1.1.7 in the UK (credit @COG-CONSORTIUM) in November 2020 <https://vimeo.com/493031746>



# Re-designing a visualisation tool

Covid19 dashboards



# Nextstrain

DOCS HELP LOGIN

**Dataset**

groups

neherlab

ncov

united-kingdom

**Date Range**

2019-12-20 2023-02-16

PLAY RESET

**Color By**

Clade

**Filter Data**

Type filter query here...

Currently selected filter categories:

1 x Country

**Tree Options**

Layout

RECTANGULAR

RADIAL

UNROOTED

<https://nextstrain.org>

## SARS-CoV-2 phylogeny

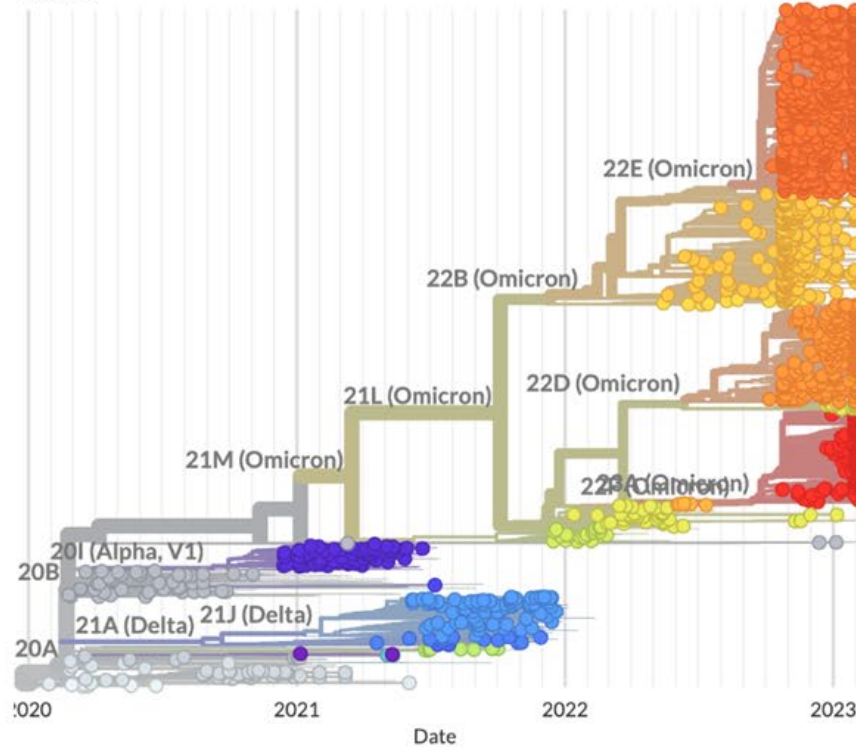


Built with [neherlab/ncov-simple](#). Maintained by [Cornelius Roemer](#) and [Richard Neher](#). Enabled by data from [GISAID](#).

Showing 1955 of 3881 genomes sampled between Jan 2020 and Feb 2023. Filtered to [United Kingdom \(1955\)](#)

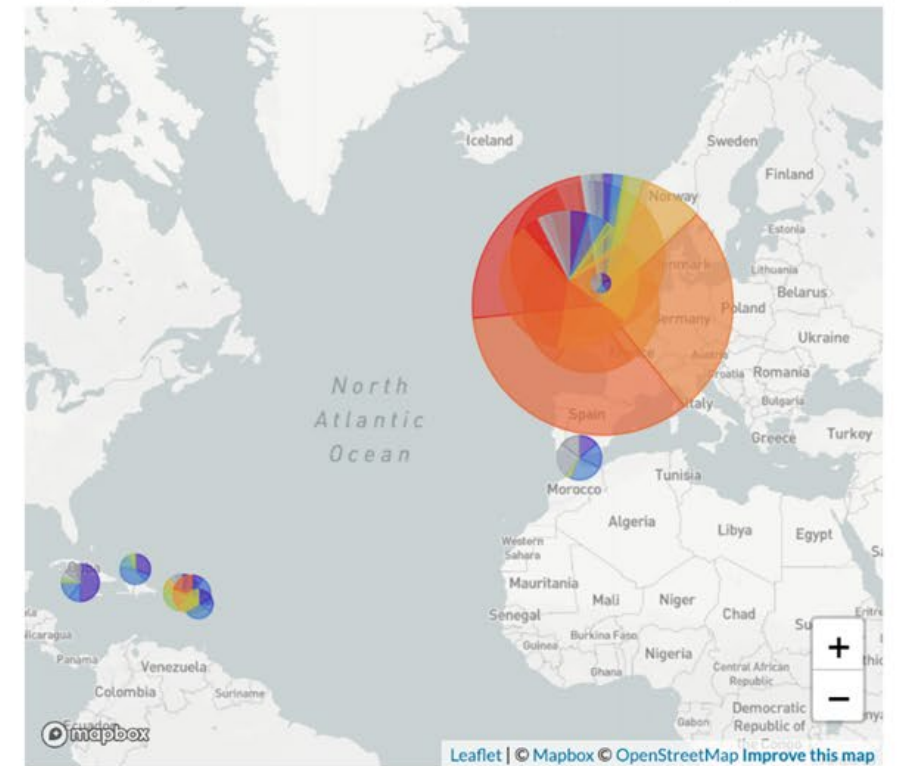
### Phylogeny

Clade



### Geography

RESET ZOOM



### Diversity

