

Forecasting with Generalised Additive Models (GAMs) in R

Forecasting for Social Good

21 February 2024

Dr Nicola Rennie

Welcome!

About me

Lecturer in Health Data Science at
Lancaster Medical School.

Academic background in statistics, with
experience in data science consultancy.

Blog about R and data science at
nrennie.rbind.io/blog.



What to expect during this workshop

The workshop will run for *2 hours*.

- Combines slides, live coding examples, and exercises for you to participate in.
- Ask questions in the chat throughout!

What to expect during this workshop

I hope you end up with more questions than answers after this workshop!



Source: [giphy.com](#)

What you need for this workshop

- You are comfortable with simple operations in R.
- You know how to perform linear regression.

Workshop resources

GitHub: github.com/nrennie/f4sg-gams

Slides: nrennie.github.io/f4sg-gams

Data

We'll be modelling Covid-19 data during the workshop. The data is from the {COVID19} R package.

- For Great Britain and France between January 2021 and January 2022.
- For each day, data on cases, tests, vaccines, deaths, ...

Guidotti, E., Ardia, D., (2020), "COVID-19 Data Hub", Journal of Open Source Software 5(51):2376, doi: [10.21105/joss.02376](https://doi.org/10.21105/joss.02376).

Data

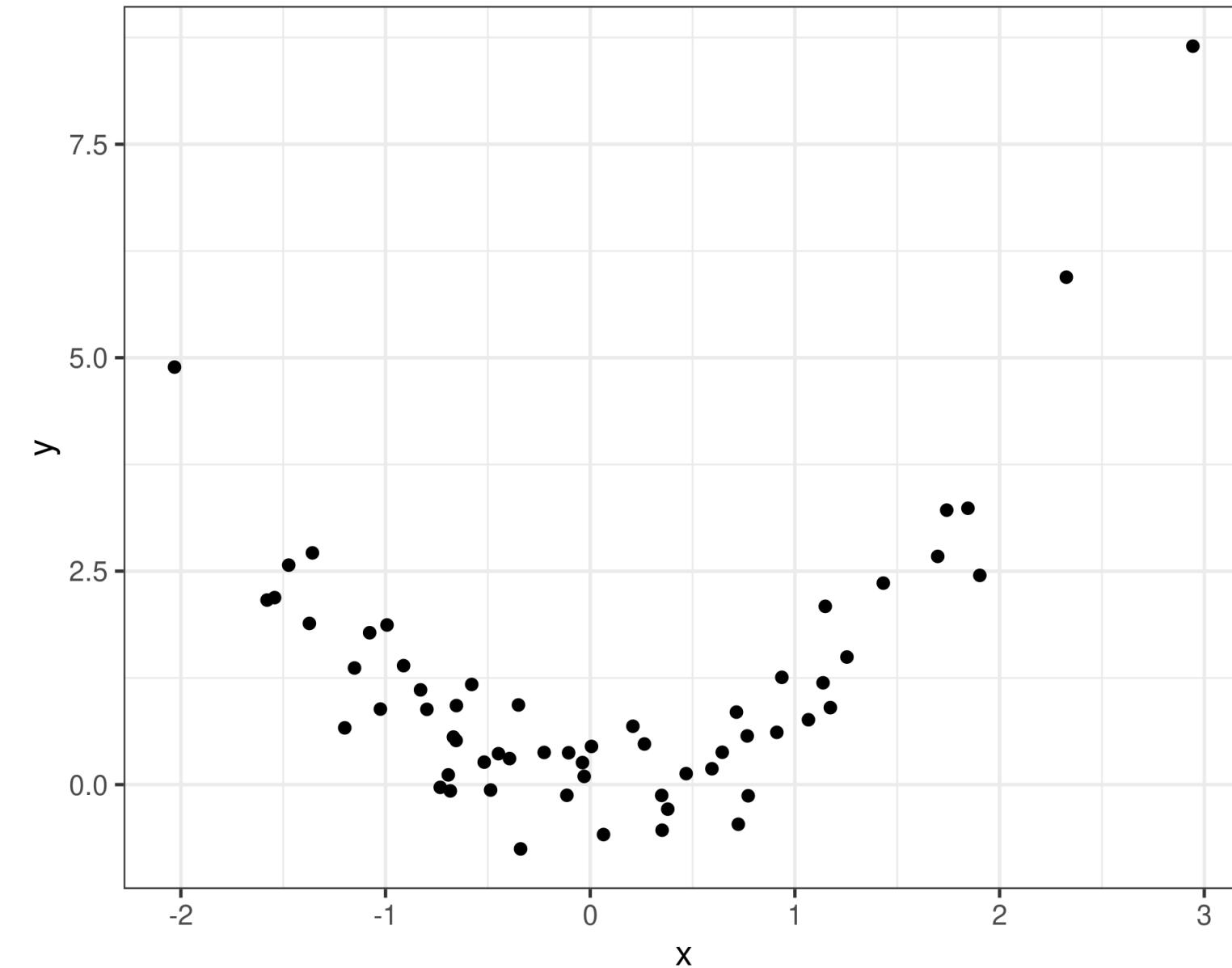
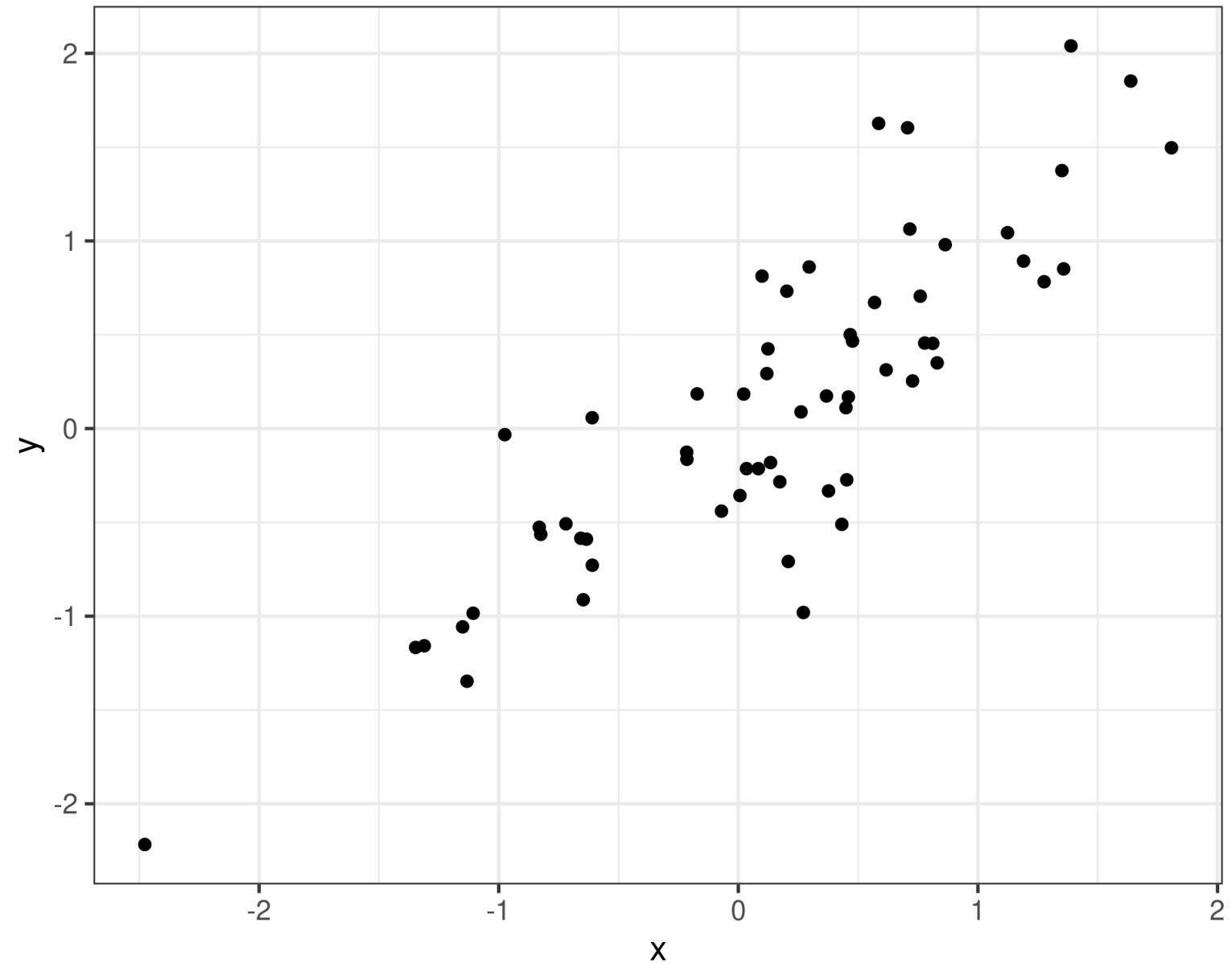
How to download:

```
1 covid <- readr::read_csv(  
2   "https://raw.githubusercontent.com/nrennie/f4sg-gams/main/data/covid.csv"  
3 )
```

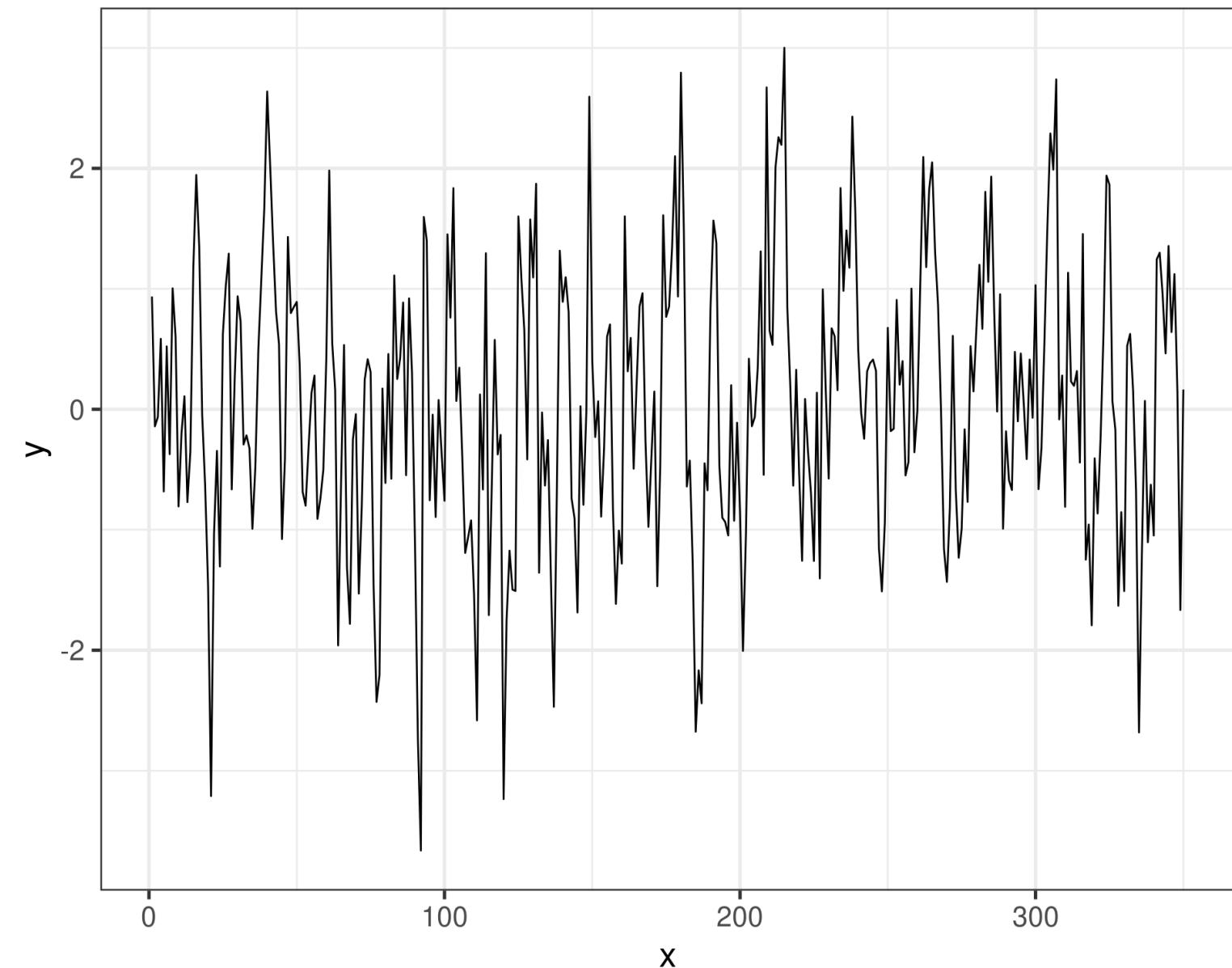
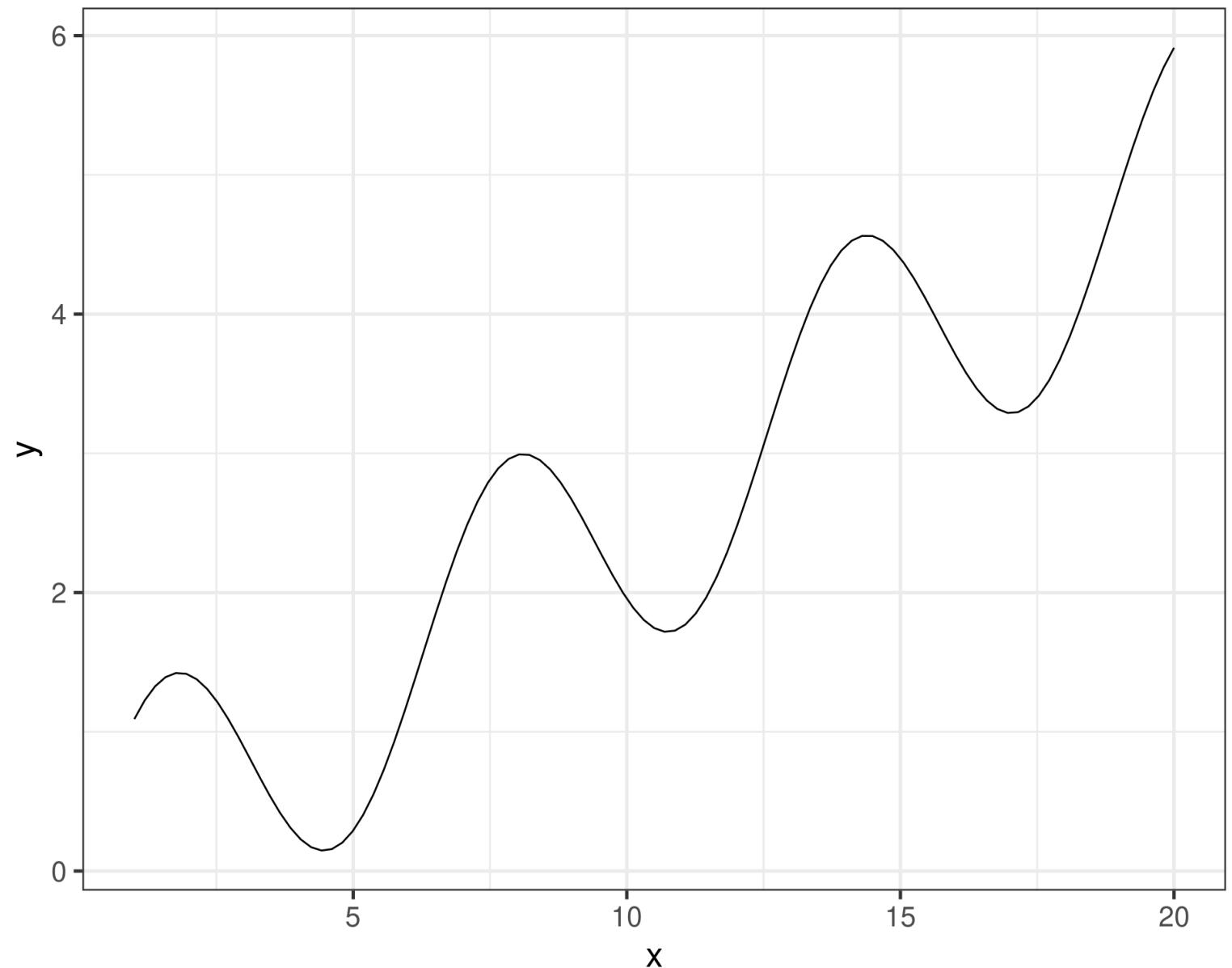
- See **data/** folder on GitHub for pre-processing.

What are generalised additive models?

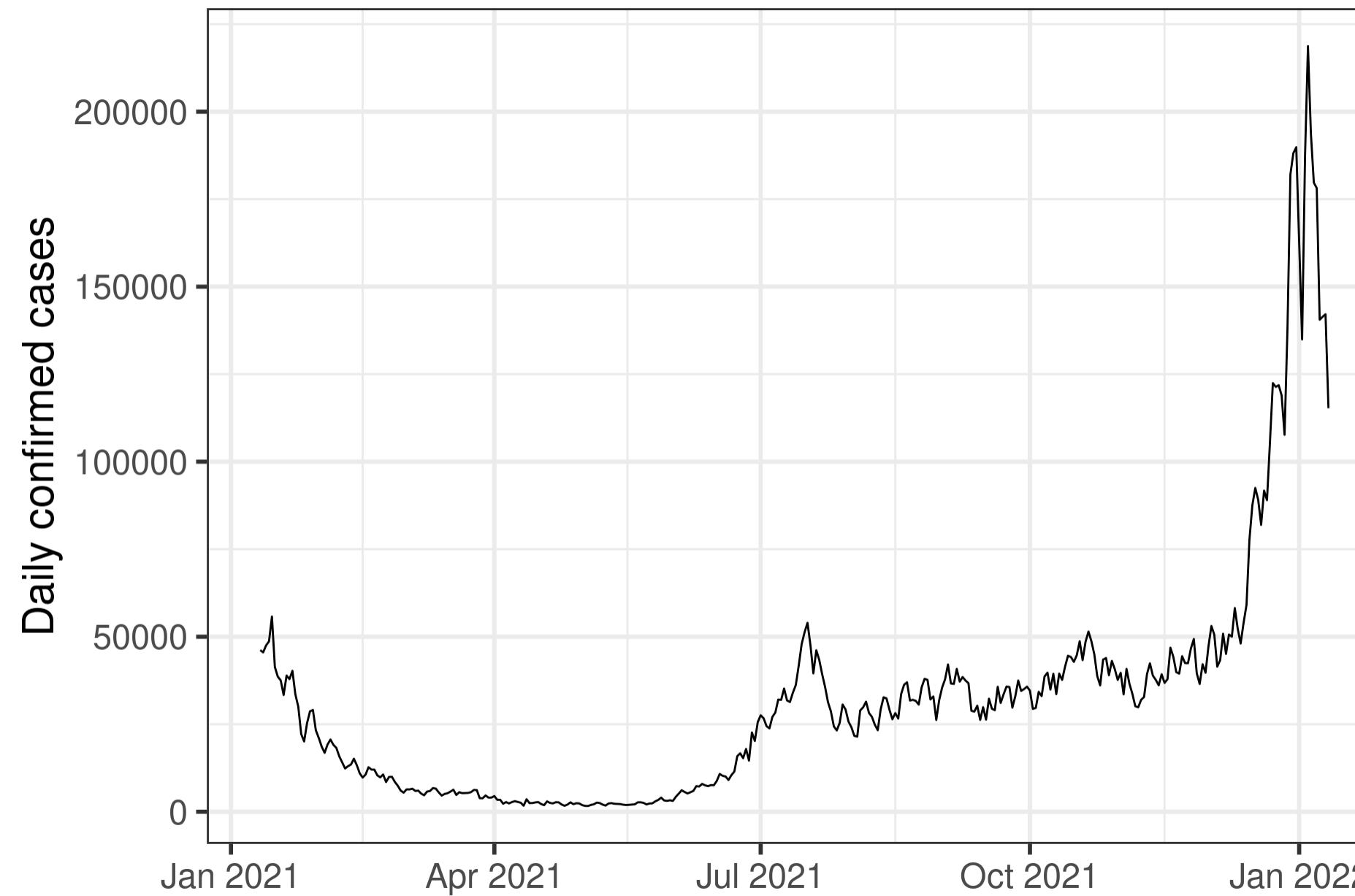
Sometimes relationships are easy to spot...



Sometimes relationships are easy to spot...



Other times not so much!



What are generalised additive models?

Let's start with linear models...

$$\mathbb{E}(Y) = \alpha + x_1 + x_2 + \dots + x_p$$

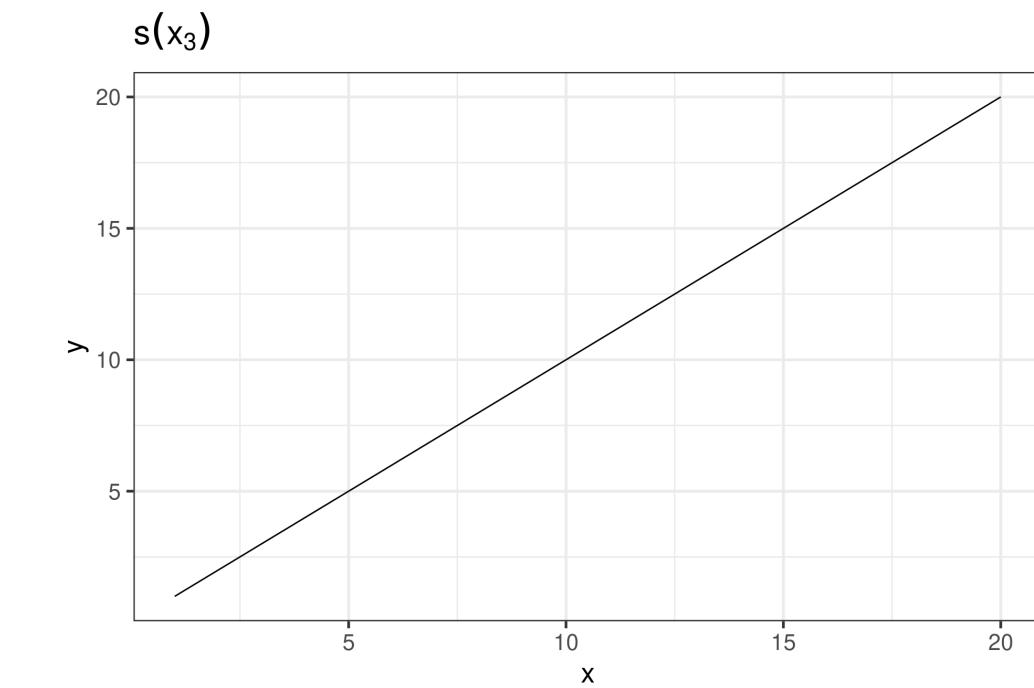
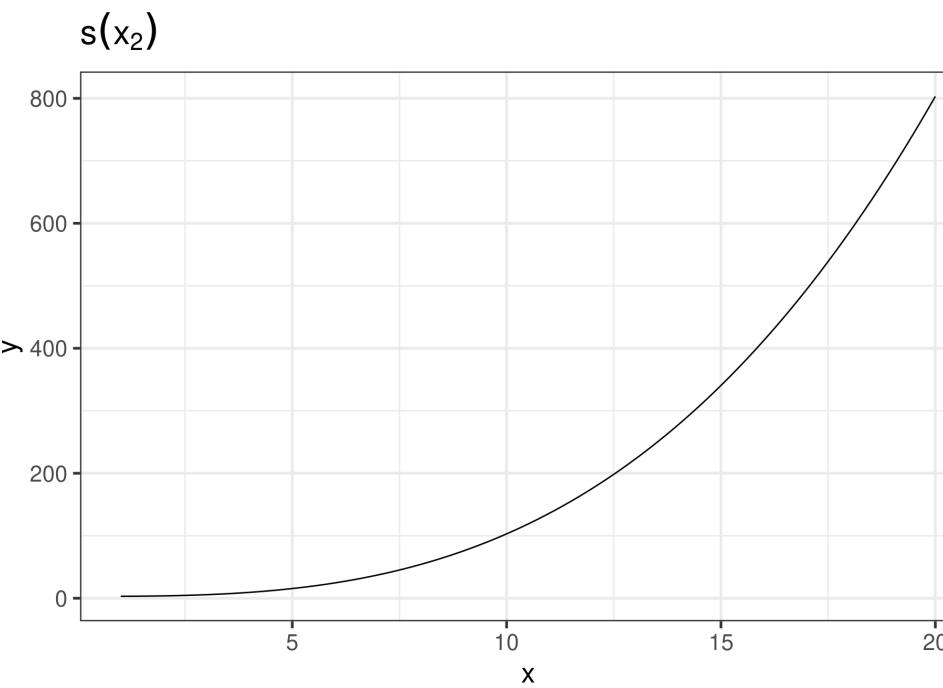
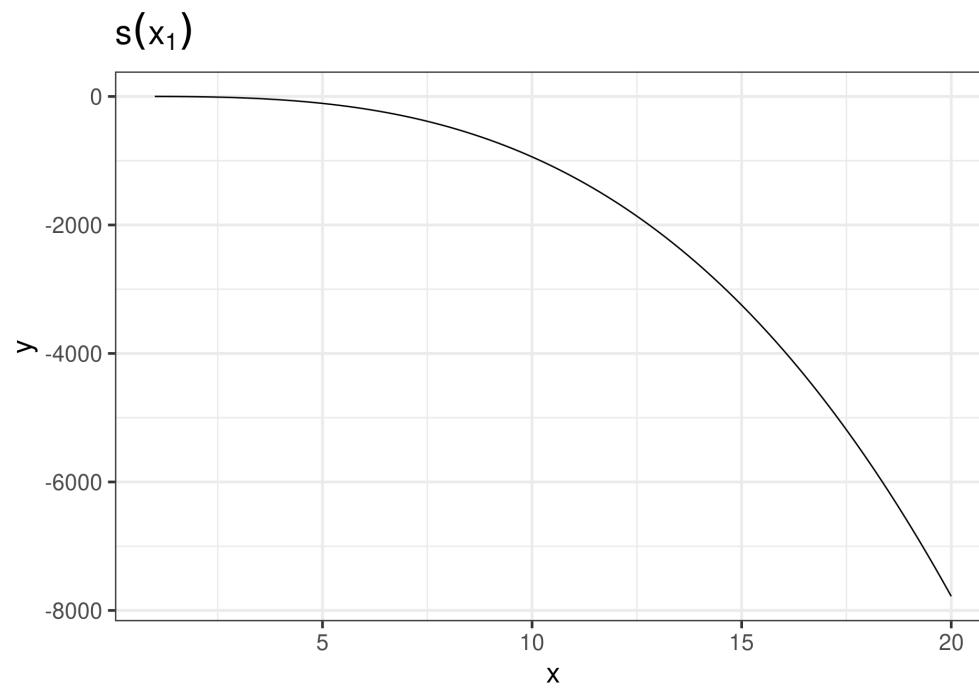
Then generalised linear models...

$$g(\mathbb{E}(Y)) = \alpha + x_1 + x_2 + \dots + x_p$$

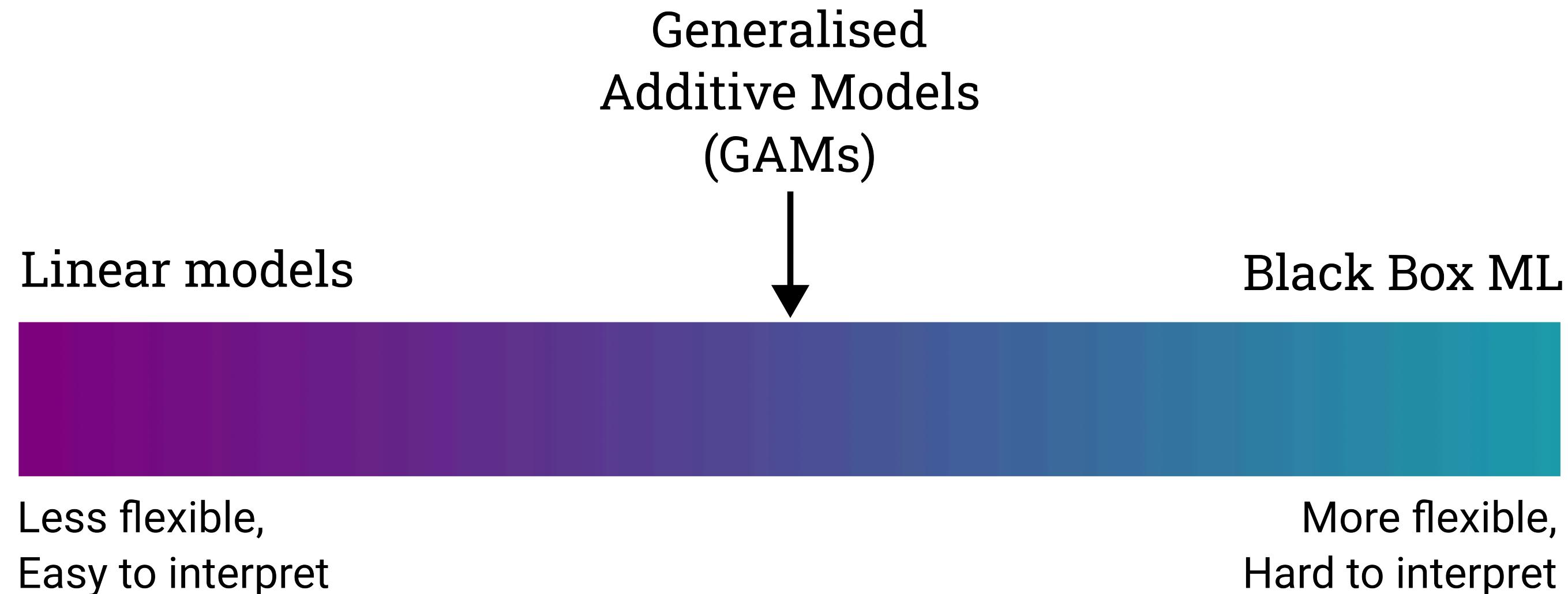
What are generalised additive models?

Generalised additive models are essentially just a sum of smooth functions of the explanatory variables.

$$g(\mathbb{E}(Y)) = \alpha + s(x_1) + s(x_2) + \dots + s(x_p)$$



What are generalised additive models?



Generalised additive models in R

There are several packages for fittings GAMs in R. The two main packages are:

- {gam}
- {mgcv}

We'll be using {mgcv}.

| Hint: don't load both packages at the same time!

Live demo!

See `examples/example_01.R` for full code.

Exercise 1

Open `exercises/exercise_01.R` for prompts.

10:00

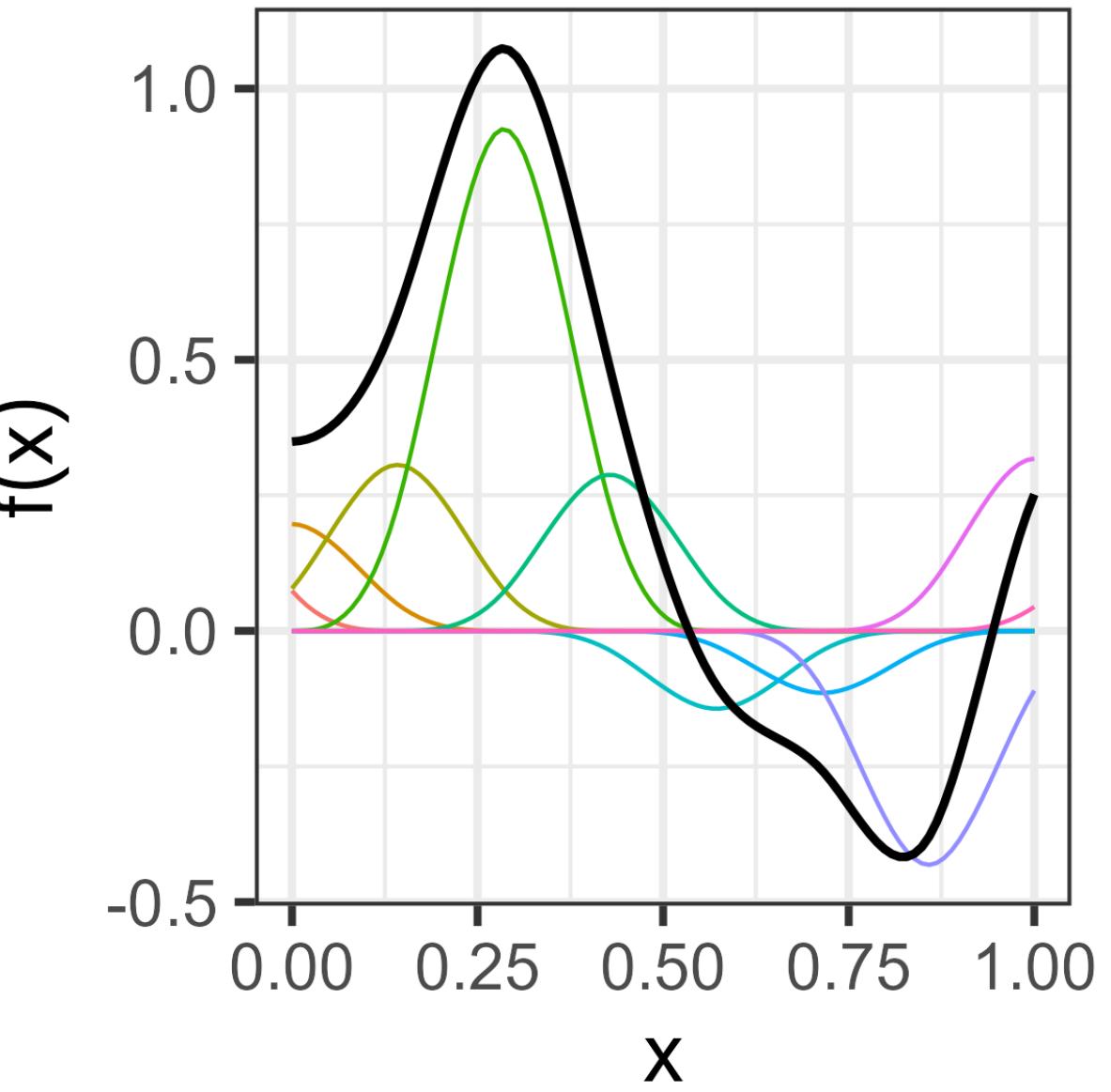
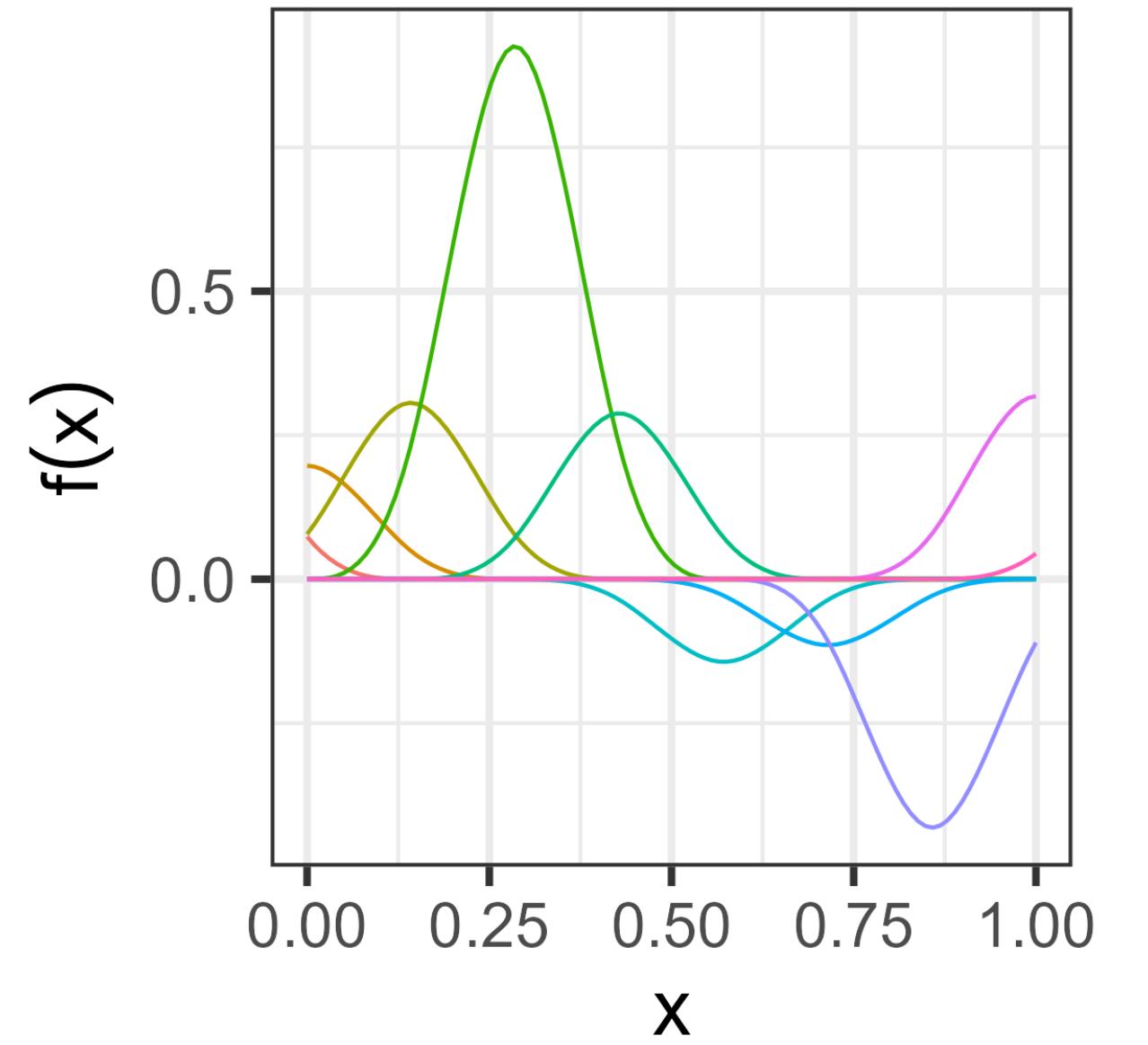
- Load the packages required for this workshop.
- Load the data for the exercises. Subset it to look at France only.
- Split the data into training and testing sets.
- Create a plot of confirmed cases over time.
- Fit a linear model using the `gam()` function in `{mgcv}` with `confirmed` as the outcome, and `date_obs` as the explanatory variable.

See `exercise_solutions/exercise_solutions_01.R` for full solution.

Generalised additive models in R

Fitting smooth functions

We want to fit some sort of smooth function to each predictor. We can add together *basis* functions.



Fitting smooth functions

```
1 gam_0 <- gam(confirmed ~ date_obs, data = gbr_train)
2 gam_1 <- gam(confirmed ~ s(date_obs), data = gbr_train)
```

There are different smooth classes available: `s()`, `te()`, `ti()`, and `t2()`. Smooth classes are invoked directly by `s()` terms, or as building blocks for tensor product smoothing via `te()`, `ti()` or `t2()` terms.

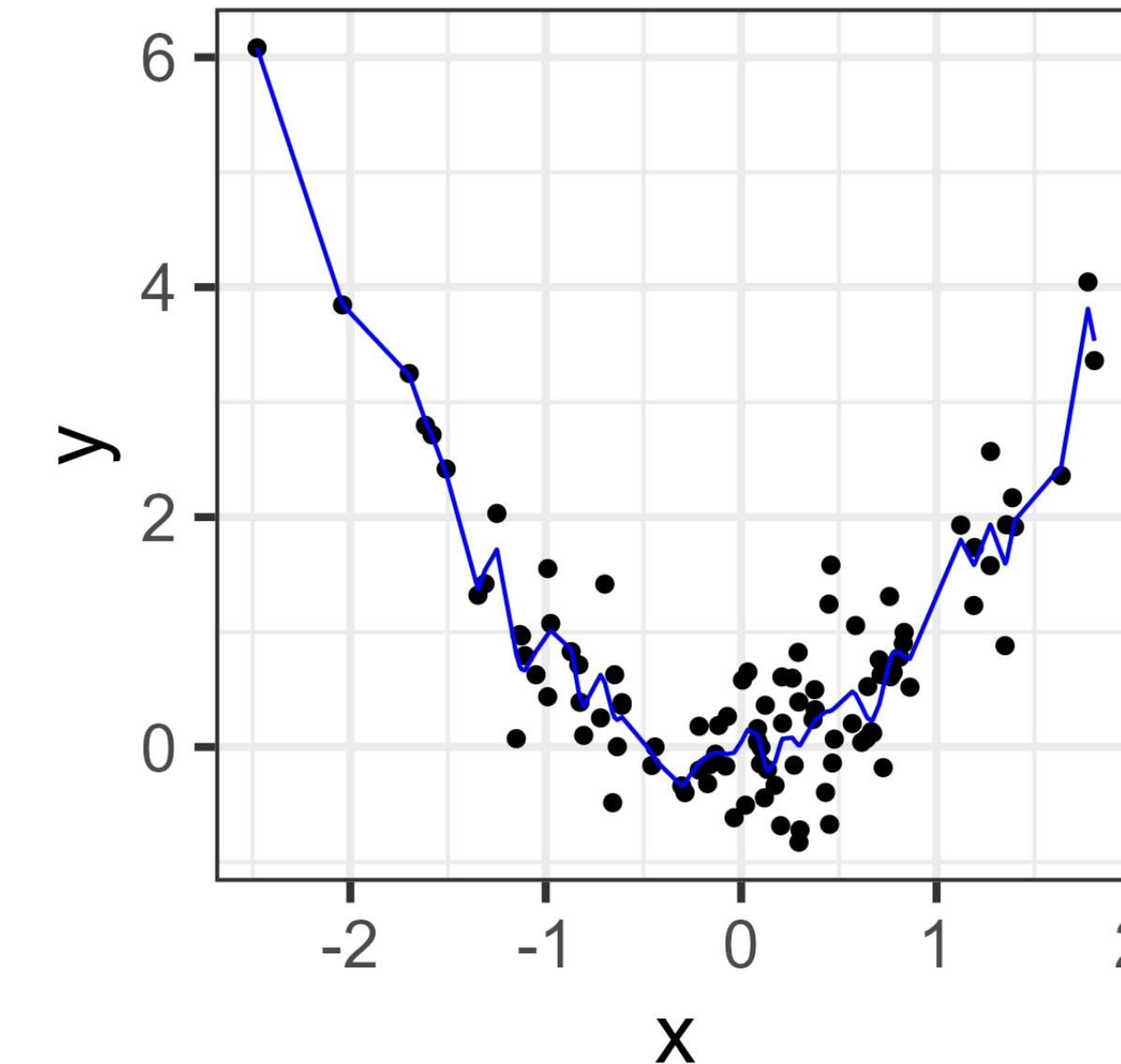
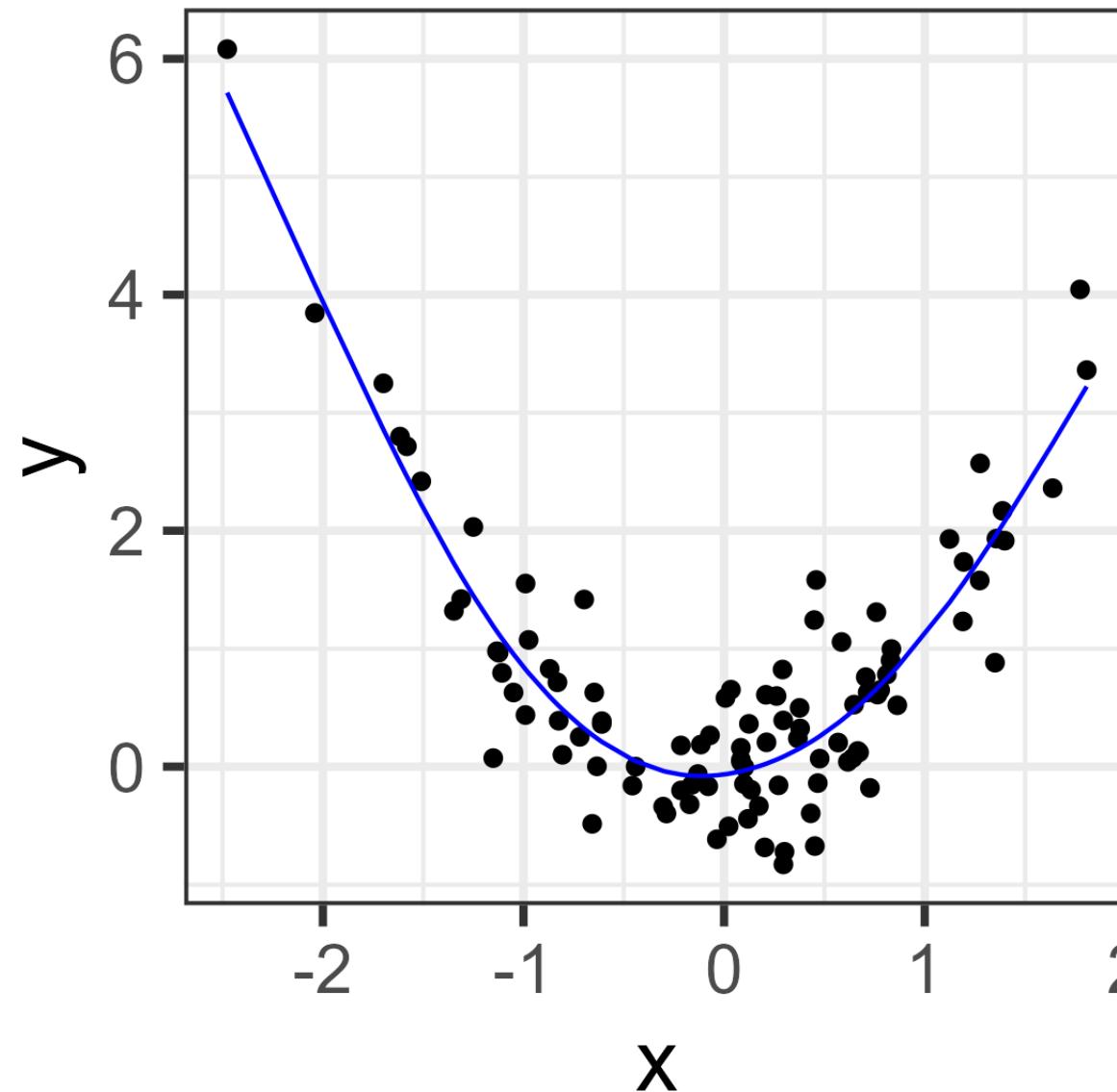
There are also different types of splines, specified by the `bs` argument. See `?smooth.terms` for details.

Multiple terms

```
1 fit <- gam(y ~ s(x1) + s(x2) + x3, data = data)
```

- You can include multiple predictors in `gam()`.
- You can use different types of smoothing for each one.
- You can include non-smooth terms (esp. categorical) variables as well.

How smooth is smooth enough?



Additional arguments

- Level of smoothing can be controlled with `sp` in `gam()` or `s()`.
- You can choose the dimension of the basis function using `k` in `s()`.

Hint: look at `?gam` and `?s` to see the wide range of arguments.

Live demo!

See `examples/example_02.R` for full code.

Exercise 2

Open `exercises/exercise_02.R` for prompts.

10:00

- Fit a GAM using `s()` and `gam()` to the confirmed cases in France.
- Try adding different additional terms.
- Try varying the smoothing parameter `sp` for each variable.
- Plot the smooth functions for the predictors, and the outcome.

See `exercise_solutions/exercise_solutions_02.R` for full solution.

Predicting with GAMs

After you've fitted a GAM...

- Is it a reasonable model?
- Is it better than other models?
- How well does it fit to the data?
- What are the forecasted values?
- How do we interpret the model?

Live demo!

See `examples/example_03.R` for full code.

Exercise 3

Open `exercises/exercise_03.R` for prompts.

10:00

- Run `gam.check()` on your GAM from Exercise 2. Do you need to refit?
- Obtain the fitted values.
- Make a forecast for the range of the test data.
- Inspect the contribution of different terms.

See `exercise_solutions/exercise_solutions_03.R` for full solution.

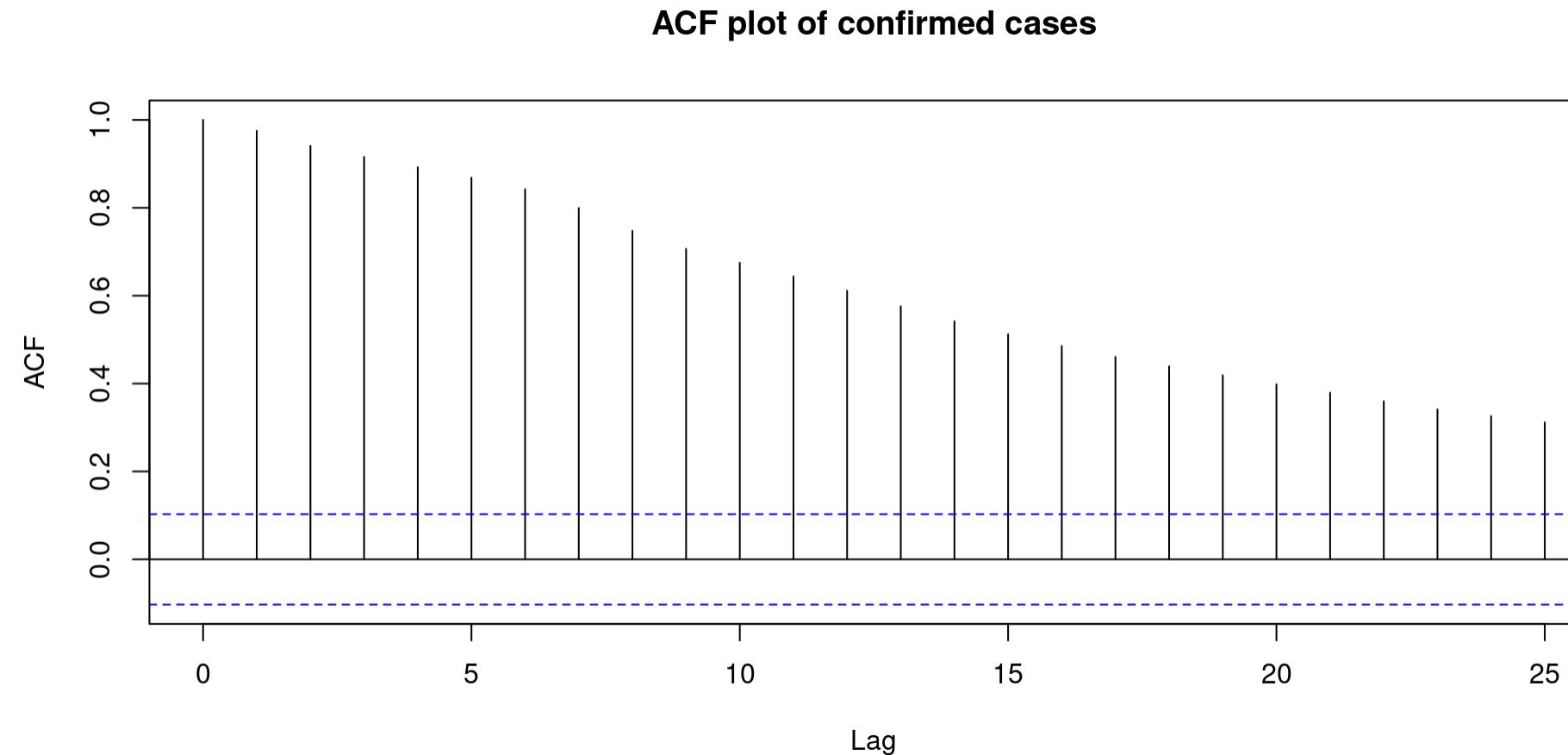
When GAMs don't quite work...

Common problems

- Unstable estimates: when a smooth term can be approximated by some combination of the others.
- Computationally expensive for large data sets: see `bam()`.
- Lack of independence between observations: see `gamm()`.
- Unstable predictions outside the range of training data: see `{mgcv}`

Are observations independent?

```
1 acf(gbr_data$confirmed, main = "ACF plot of confirmed cases")
```



Generalized additive mixed effect models (GAMMs)

- Like GAMs, GAMMs allow for non-linear relationships between predictors and the response variable by fitting smooth functions to each predictor.
- GAMMs also allow for the inclusion of random effects, which capture the variability of observations within groups or clusters.
- This includes adding correlation structures.

Live demo!

See `examples/example_04.R` for full code.

Exercise 4

Open `exercises/exercise_04.R` for prompts.

10:00

- Look at the ACF and PACF plots of the residuals from your previous GAM.
- Try fitting a GAMM model instead.

See `exercise_solutions/exercise_solutions_04.R` for full solution.

Additional resources

- GAMs resource list: github.com/noamross/gam-resources
- {mgcv} course: noamross.github.io/gams-in-r-course
- GAMMs: r.qcbs.ca/workshop08/book-en/introduction-to-generalized-additive-mixed-models-gamms.html
- DGAMs for forecasting: doi.org/10.1111/2041-210X.13974

Useful links

GitHub: github.com/nrennie/f4sg-gams

Slides: nrennie.github.io/f4sg-gams

Posit Cloud: posit.cloud/content/7637971

 [nicola-rennie](#)

 [@fosstodon.org/@nrennie](#)

 [nrennie](#)

 [nrennie.rbind.io](#)

 [chicas.lancaster-university.uk](#)