

Week 11: Splines

02/04/23

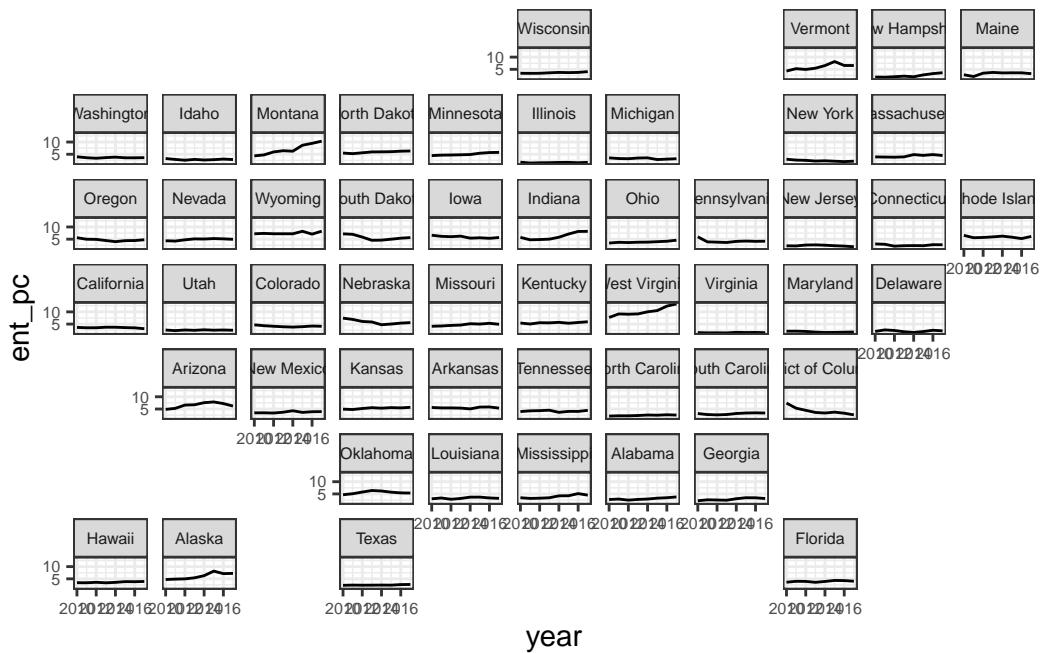
Overview

In this lab you'll be fitting a second-order P-Splines regression model to foster care entries by state in the US, projecting out to 2030.

Here's the data

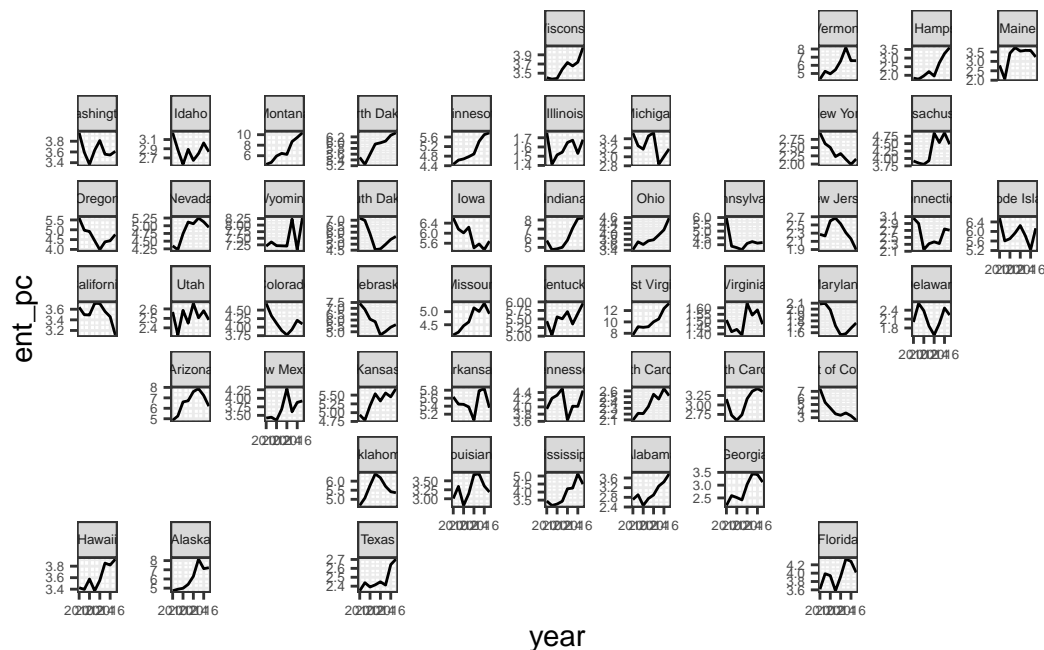
Question 1

Make a plot highlighting trends over time by state. Might be a good opportunity to use `geofacet`. Describe what you see in a couple of sentences.



From the plot above, the scale of changes are quite different, i.e. while some states (such as Most, West Virginia) observed significant change in the proportion of foster care entries, the change is small for most states.

To see the change more visible for each states, the following plot gives one with the scale adjusted for each state. From this plot, you can see the trend of the proportion of foster care entries over time varies a lot from state to state. You see some states observed the increase while others decreased or going up and down, etc.



Question 2

Fit a hierarchical second-order P-Splines regression model to estimate the (logged) entries per capita over the period 2010-2017. The model you want to fit is

$$\begin{aligned}
 y_{st} &\sim N(\log \lambda_{st}, \sigma_{y,s}^2) \\
 \log \lambda_{st} &= \alpha_k B_k(t) \\
 \Delta^2 \alpha_k &\sim N(0, \sigma_{\alpha,s}^2) \\
 \log \sigma_{\alpha,s} &\sim N(\mu_\sigma, \tau^2)
 \end{aligned}$$

Where $y_{s,t}$ is the logged entries per capita for state s in year t . Use cubic splines that have knots 2.5 years apart and are a constant shape at the boundaries. Put standard normal priors on standard deviations and hyperparameters.

```

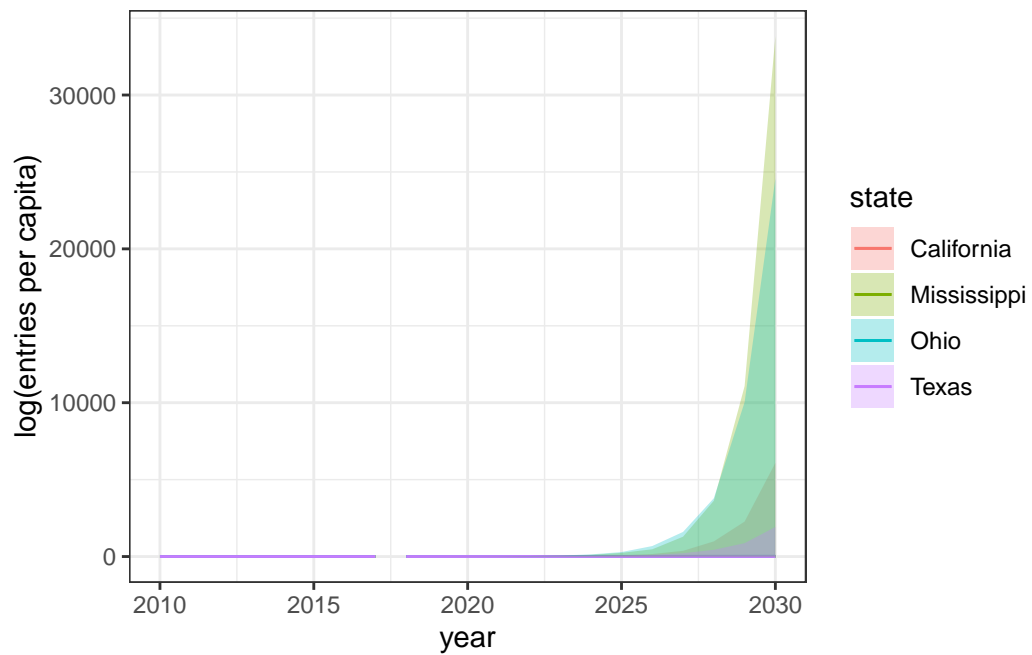
Running /Library/Frameworks/R.framework/Resources/bin/R CMD SHLIB foo.c
clang -mmacosx-version-min=10.13 -I"/Library/Frameworks/R.framework/Resources/include" -DNDEBUG
In file included from <built-in>:1:
In file included from /Library/Frameworks/R.framework/Versions/4.2/Resources/library/StanHeaders/include/stan/math/
In file included from /Library/Frameworks/R.framework/Versions/4.2/Resources/library/RcppEigen/include/Eigen/src/Core/
In file included from /Library/Frameworks/R.framework/Versions/4.2/Resources/library/RcppEigen/include/Eigen/src/Core/
namespace Eigen {
~

/Library/Frameworks/R.framework/Versions/4.2/Resources/library/RcppEigen/include/Eigen/src/Core/
namespace Eigen {
~
;
In file included from <built-in>:1:
In file included from /Library/Frameworks/R.framework/Versions/4.2/Resources/library/StanHeaders/include/stan/math/
In file included from /Library/Frameworks/R.framework/Versions/4.2/Resources/library/RcppEigen/include/Eigen/Core:
#include <complex>
~~~~~~
3 errors generated.
make: *** [foo.o] Error 1

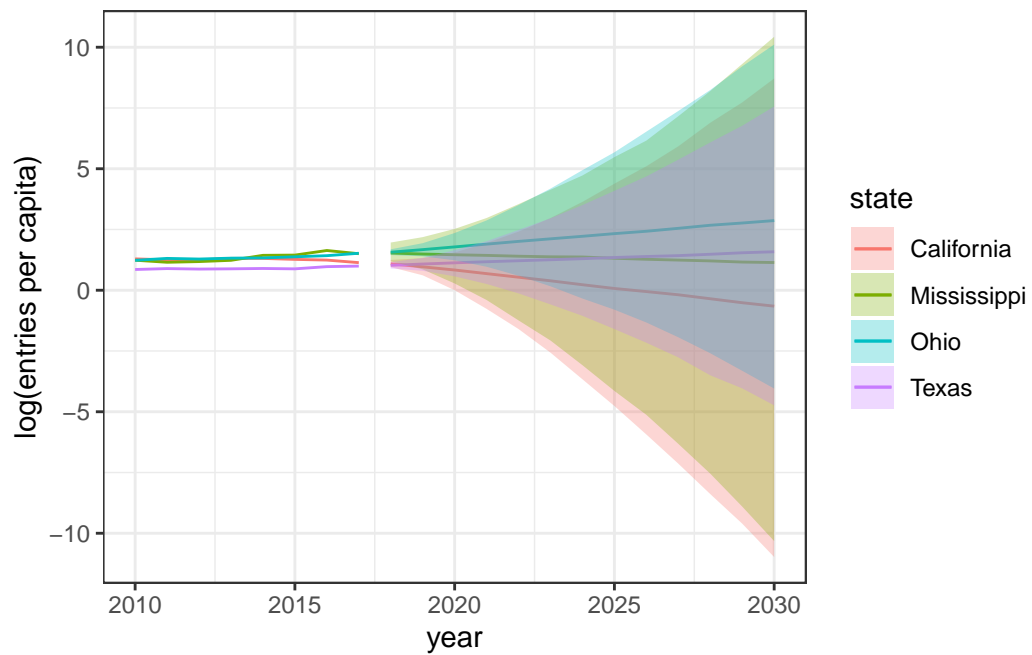
```

Question 3

Project forward entries per capita to 2030. Pick 4 states and plot the results (with 95% CIs).
Note the code to do this in R is in the lecture slides.



It is hard to see how well the projection is with this scale. So we take $\log(\text{ent_pc})$ instead.



Question 4 (bonus)

P-Splines are quite useful in structural time series models, when you are using a model of the form

$$f(y_t) = \text{systematic part} + \text{time-specific deviations}$$

where the systematic part is model with a set of covariates for example, and P-splines are used to smooth data-driven deviations over time. Consider adding covariates to the model you ran above. What are some potential issues that may happen in estimation? Can you think of an additional constraint to add to the model that would overcome these issues?