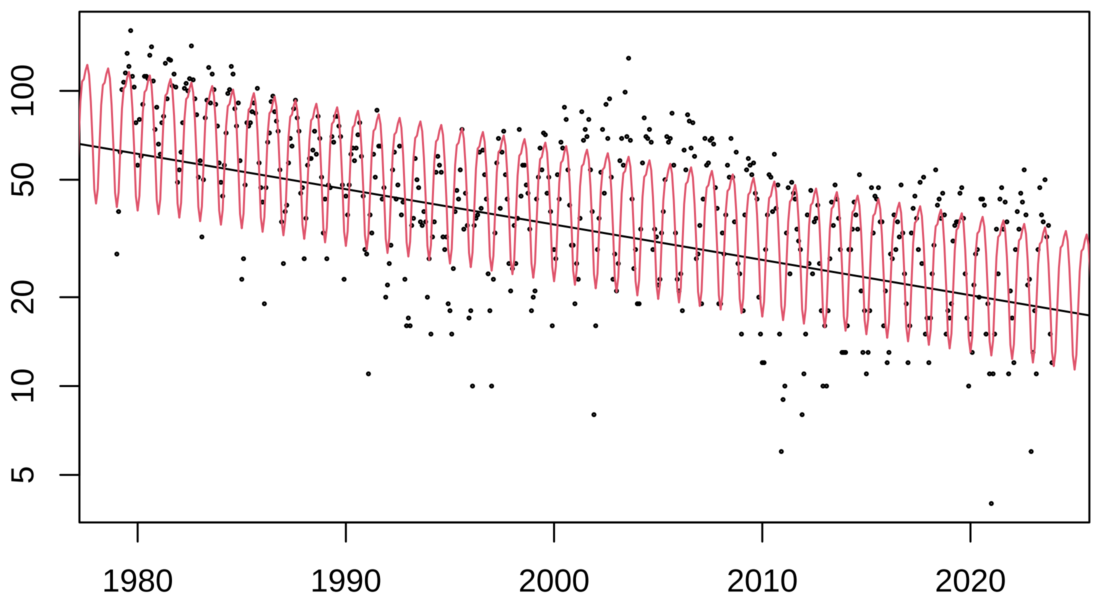
STA303 Assignment 2

Talia Fabregas Kasandra Tworzynaski

March 24, 2025

Q1: Motorcycle deaths




```
##         date killed serious slight dateInt logMonthDays month
## 1 1979-01-01     28     737   1677    3287     3.433987   Jan
## 2 1979-02-01     39     854   1914    3318     3.332205   Feb
## 3 1979-03-01     62    1302   2895    3346     3.433987   Mar
## 4 1979-04-01    101    1540   3406    3377     3.401197   Apr
## 5 1979-05-01    107    1774   4093    3407     3.433987   May
## 6 1979-06-01    115    1912   4441    3438     3.401197   Jun
```

1. Write down, in equations not R code, a generalized additive model suitable for this problem. Explain each of the parts of the model and give a rationale for them (i.e. "The response variable is Gamma distributed because the number of deaths must be positive"). (4 points)

$$Y_i|U \sim Pois(D_i\lambda_i)\log(\lambda_i) = X_i\beta + U(t_i)U(.) \sim IWP_2(\sigma)$$

where:

- Y_i = the number of motorcycle deaths in month i
- D_i = the number of days in month i
- λ_i = intensity of expected number of deaths
- X_i = month indicator variables (seasonality, month as a factor)

- β = fixed month effect
- $U(t_i)$ = smooth term, second-order Integrated Wiener Process (IWP)

The response variable, Y_i, is Poisson because the number of motorcycle deaths is a non-negative count variable. D_i is the number of deaths in month i and λ_i is the intensity. We will use D_i as an offset because exposure time varies by the length of the month (i.e. February has 28 or 29 days and January has 31 days). The log link function ensures that the predicted counts are non-negative and creates a multiplicative relationship between the predictors and response.

```
model1 <- gam(killed ~ s(dateInt) + month + offset(logMonthDays),
              data=x,
              family=poisson,
              method = 'ML')
```

2. Show R code to fit the model using the `mcgv` package

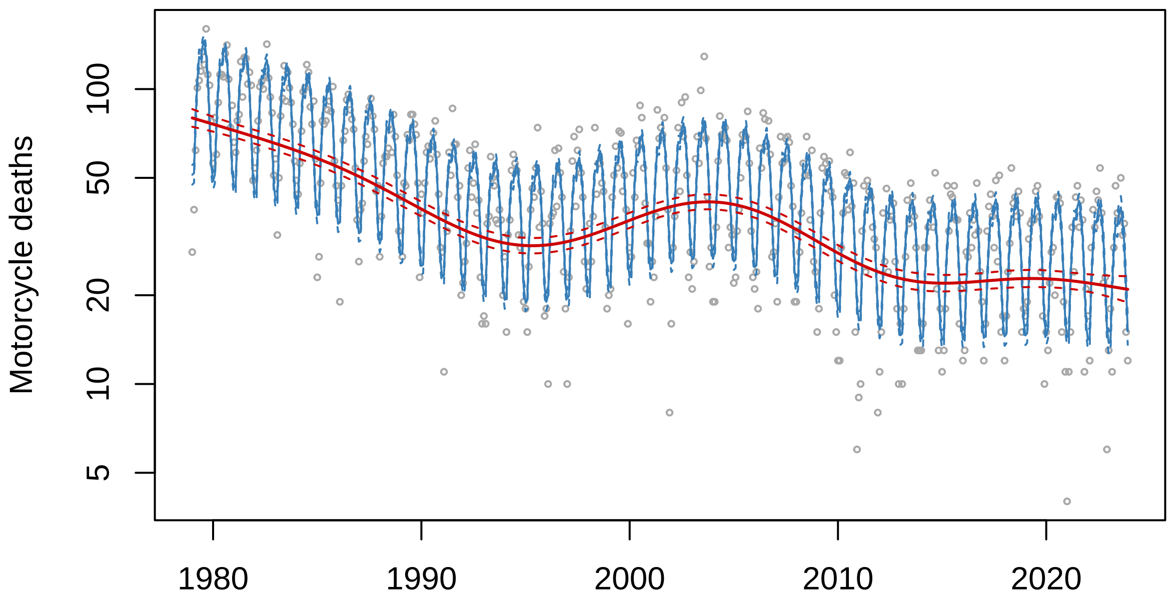


Figure 1: Motorcycle deaths have declined over time, since 1980.

3. Produce a figure similar to Figure fig. 1 which is able to visualize the trend estimated from the motorcycle data. You're marked on the figure looking professional (with clear labels and a caption) as well as conveying the important statistical information (prediction intervals as well as point predictions). (4 points)

Q2: Heat

1.

2.

3.

4.

```
##           date killed serious slight dateInt logMonthDays month
## 100 1987-04-01     57    1090   2537    6299     3.401197   Apr
```

{r. echo=FALSE, message=FALSE, warning=FALSE} x$dateInt = as.integer(x$Date) x$yearFac = factor(format(x$Date, "%Y")) # xSub = x[x$summer & !is.na(x$Max.Temp),] # res1 = gam(update.formula(Max.Temp ~ s(dateInt, pc = as.integer(as.Date("1990/7/1")), # +   k = 100) + s(yearFac, bs = "re"), Pmisc::seasonalFormula(period = 365.25, # +   harmonics = 1:2, var = "dateInt")), data = xSub, method = "ML", # +   optimizer = "efs")