Refer to the Air Quality data described previously, and the analyses we have done with Ozone as the response variable, and the five explanatory variables (including the two engineered features).

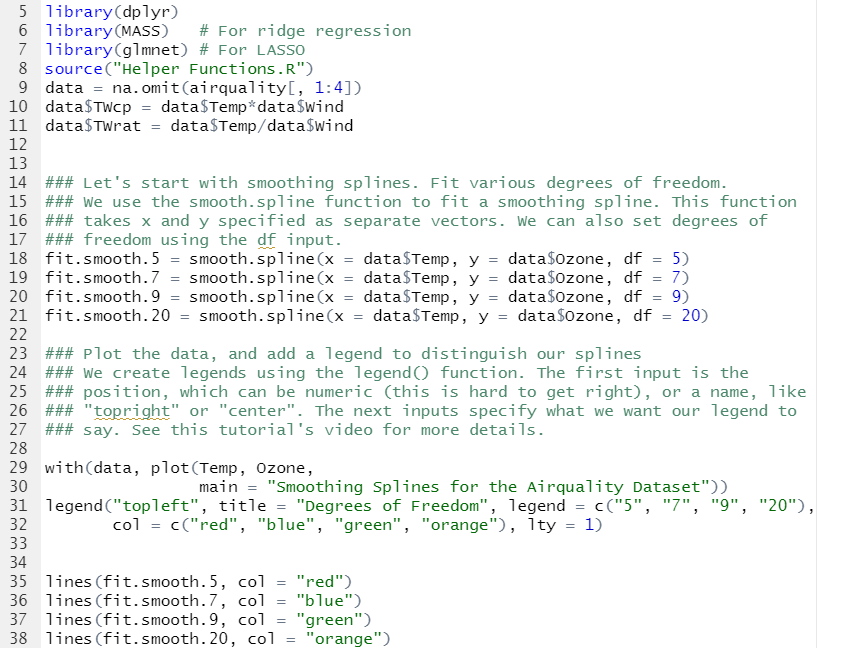
1. Use smoothing splines to model the relationship between Ozone and Temp:

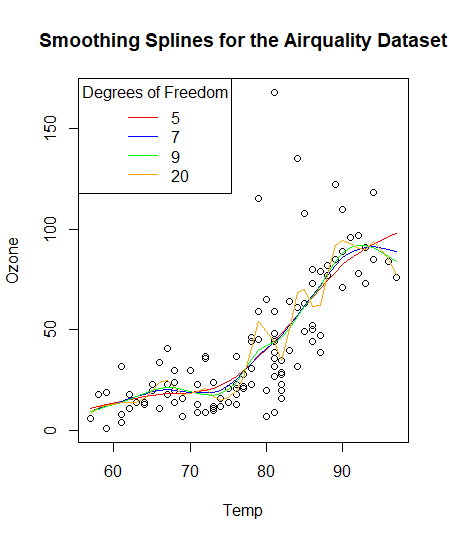
(a) On one graph, plot the data along with fits of smoothing splines with 5, 7, 9, and

20 DF.

i. **Present the plot. Be sure to add a legend and use different colours**

**for the different functions**



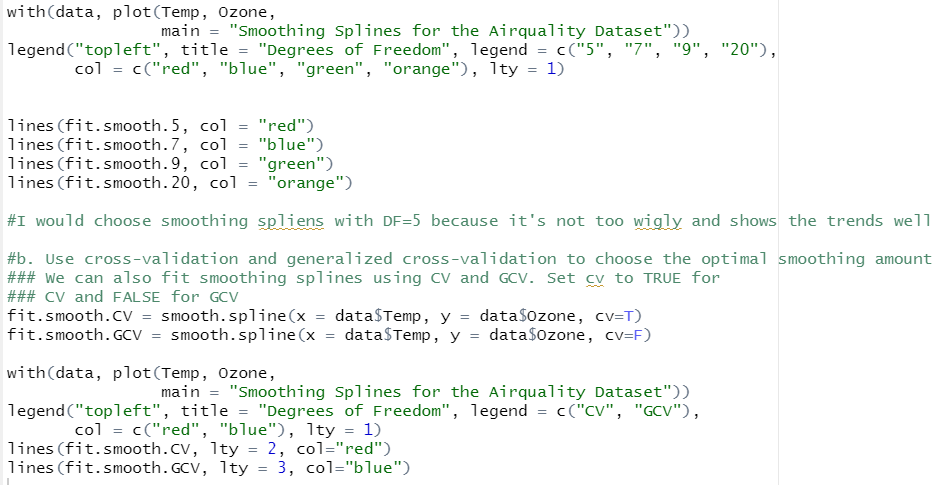


ii. If you had to choose one model, **which would it be? Why?**



(b) Use cross-validation and generalized cross-validation to choose the optimal smoothing

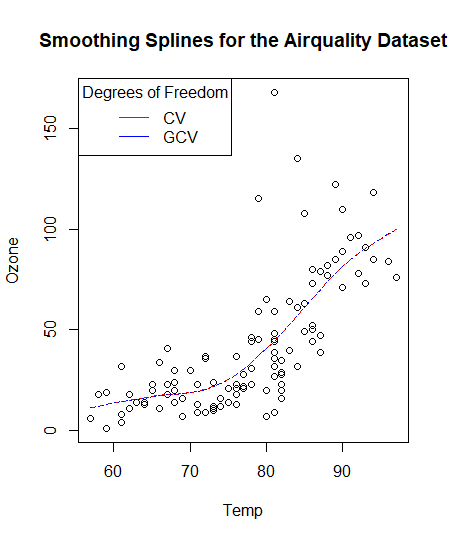
Amount



i. **How many DF does each method suggest to use?**

**->** CV suggests to use 4.588 DF and GCV suggests to use DF 4.56

ii. **Show the fits on one plot with the data.**



iii. **Comment on the quality of each fit.**

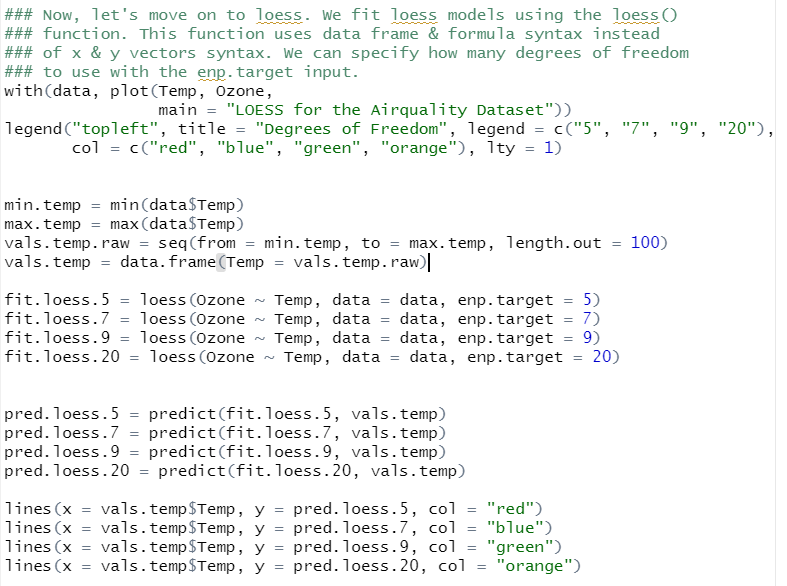
* Both are very similar and doing well in that it’s not too wiggly but fits the trend.

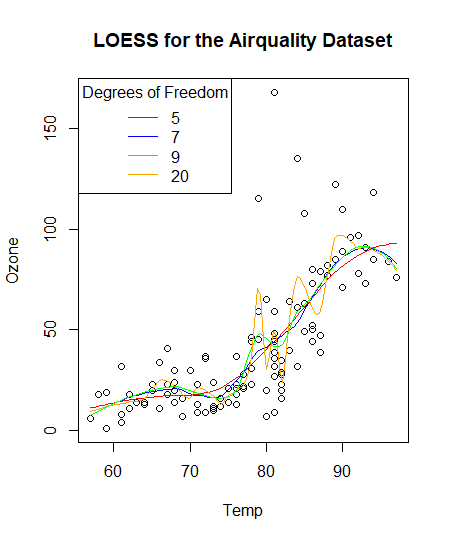
2. Repeat part (a) from Exercise 1 using LOESS.

(a) On one graph, plot the data along with fits of LOESS with 5, 7, 9, and 20 DF.

i. **Present the plot. Be sure to add a legend and use different colours**

**for the different functions**





ii. If you had to choose one model, **which would it be? Why?**

I would choose DF=5, because it’s not too wiggly and showing the trend well

fit.smooth.CV = smooth.spline(x = data$Temp, y = data$Ozone, cv=T)

fit.smooth.GCV = smooth.spline(x = data$Temp, y = data$Ozone, cv=F)