Complete these problems in a well-written RMarkdown document and upload the corresponding PDF or html file.

From the textbook, do problems **2.4 #4, 7, 9 (**5 points each).

The last problem focuses on using the Shiny app at [https://keeganhines.shinyapps.io/bias\_variance/Links to an external site.](https://keeganhines.shinyapps.io/bias_variance/). Before working on this problem, load the app, read the explanation, play with the slider and the "Generate New Data" button, and answer the questions at the bottom of the page ("Check your understanding") for yourself or discuss them with others.

This problem is worth 10 points (2 points for each part).

*Model complexity = degree of the polynomial that is being fitted.*

a) Make 10 different simulations with model complexity = 1. Compute the average Residual SSE. Also find the approximate range of the highest order coefficient for these 10 simulations. This is a measure for the baseline variance for a low complexity model.

b) Make 10 different simulations with model complexity = 3. Compute the average Residual SSE. Which coefficient has the largest range in this case? What is that range? This is a measure for the variance for a medium complexity model.

c) Repeat this for model complexity = 15. Which coefficient has the largest range for these 10 simulations? What is that range?

d) How do your results illustrate the bias – variance trade-off? The answer should be a short paragraph.

e) For which model complexity do you typically obtain a curve which is most similar and overall close to the unknown curve that is to be estimated? Try multiple simulation for several different model complexities, summarize what you see, and explain your answer. Pictures or numerical results are not required.