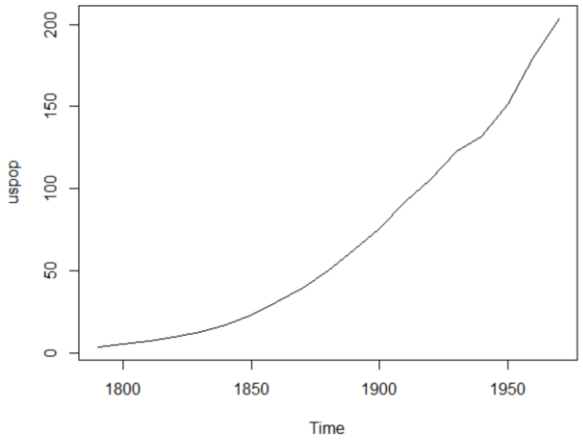
MTH 325, Quiz #11, Spring 2023 Name _____

Instructions: Answer each question as thoroughly as possible. Round answers to 4 decimal places as needed. Exact answers are best when possible. Be sure to answer all parts of each question.

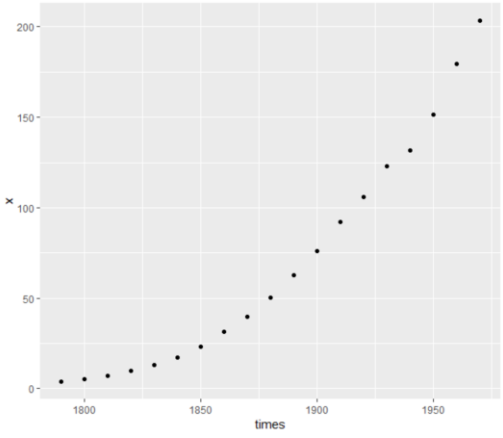
1. For this problem, you'll need to install the {xts} package. Follow the steps to convert the time series uspop (a built-in dataset in R) to a dataframe. Then perform the indicated analysis.

```
data1 <- as.data.frame(uspop)
data1$times <- rownames(data1)
data1$times<-as.numeric(data1$times)
data1$times<-(data1$times)*10+1780
data1
```

   a. Plot the data in ggplot. Paste the graph below. Describe the trend.

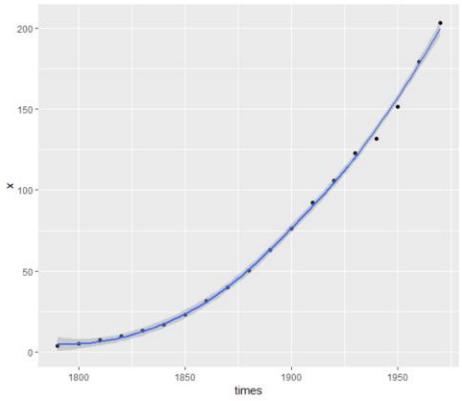


Base plot:

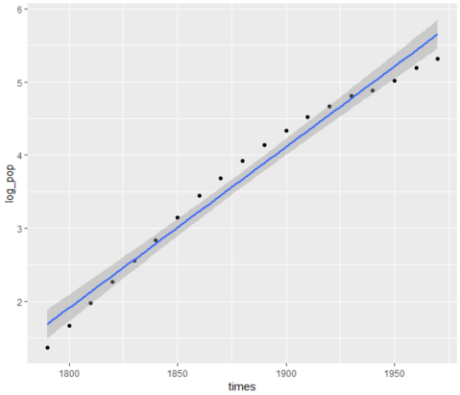


ggplot:

   b. Use geom_smooth() to plot a LOESS model to the data. Paste the graph below.



c. Create a new column of the log of the population. And replot with a linear model.



d. Write the equation of the resulting model. Perform algebra to convert this into an exponential model.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.774e+01  1.661e+00  -22.73 3.68e-14 ***
times        2.202e-02  8.829e-04   24.95 7.87e-15 ***
```

$$\ln(pop) = -37.74 + 0.02202(time)$$

$$pop = e^{-37.74+0.02202(time)}$$

$$pop = e^{-37.74}e^{0.02202(time)}$$

$$pop = (4.07 \times 10^{-17})e^{0.02202(time)}$$

The coefficient out front is small because the times in years are large.

e. Compare the results (keep in mind that the log operation will impact the size of the residual standard error). Which model appears to match the data most closely? Why? Which model do you prefer and under what circumstances?

The LOESS model visually appears to be a better fit, and the exponential model is a little nasty with the scaling factor. The exponential model could be simplified by using the census counts rather than the

year value. The exponential model has an R^2 of around 97%, so it's still a good model, and this would be a better model if you wanted to be able to interpret the model. If you were only using the model to impute a missing value, however, I would go with the loess model since the interpretation is less important there.