Data Challenge #4
Data Science I

Assignment: For Data Challenge #4 we will be using data on National Park visits on tidytuesday (https://github.com/rfordatascience/tidytuesday/tree/master/data/2019/2019-09-17). Please familiarize yourself with the data before you begin. You will be answering the following 3 questions:

1. Make a figure that shows the trend in the raw number of park visitors over time for 3 parks: Crater Lake, Joshua Tree, and Zion. Do the same but for the number of visitors normalized by state population for the state that contains the park.
2. There seems to be a decrease in visitors for both Crater Lake and Zion comparing 1942 to 1941 (less so for Joshua tree). Run a simulation for all three parks (like our restaurant letter grade simulation in class) where you assume the probability of visiting each of the parks in both years (1941 and 1942) are the same. Plot a distribution for the number of people who visited in 1942 from your simulations and compare it to the number of people who actually visited the park in 1942. Remember our rules about iteration! If you do something twice write a function. Use map and apply statements as well. (Hint: You may want to look at the park visits in 1000s -- i.e. divide by 1000 to ease the computational burden of your simulation) DON'T FORGET TO SET A SEED FOR REPRODUCIBILITY!
3. Summarize your findings from questions 1 and 2 in a paragraph.

A rubric for the assignment can be found below:

	Points	Points Possible	Description
Code Style		10	Is the code organized and well commented?
Submission		10	Was the data challenge submitted as an html document on Canvas? Does the html document look aesthetically pleasing? Did the homework contain a link for a GitHub repository? Did the repository contain the code for the assignment?
Question 1		20	Is the data wrangled properly? Are the plots accurate and well-made (title, good axis and legend titles that are not default variable names, no scientific notation…)?
Question 2		30	Is the simulation accurately carried out? Was the simulation written into a function? Are the results from the simulation

			communicated in an accurate and well-made plot? Is iteration used when appropriate? WAS A SEED SET FOR REPRODUCIBILITY?
Question 3		30	Are the results from Question 1 and 2 well described? Is the writing clear and accurate?
Total		**100**	