STAT 3280 Fall 2021 HW7

Due by the end of Nov 8, eastern time. Submit your homework by sending it to our GTA, Ruizhong Miao (rm9dd@virginia.edu), with the subject "STAT 3280-HW7: names", where the "names" should be replaced by your last name(s) of the group. Each group only has to submit it once. Make sure you include everyone's **name AND computing id on the first page**. **Missing any part of these will result in missing grades**. Please use a separate page for each problem. **And the answer to each problem cannot be longer than one page (with reasonable font size, line space, margins etc.).** You can explain how you did it in R by submitting your code with detailed explanations, but only include this part in an appendix. The GTA will not be guaranteed to look at your appendix, so make sure you explains things clearly in your main text. Notice that you are working on a visualization task. So, for each problem, make sure that your **plain language explanations should not exceed 1/2 of the paper in total** for each problem. The main results would be your figures. **You can use any software or packages for this homework.**

1. (10 pts) After the initial vote counting stage of the 2020 presidential elections, there were many claims about the potential fraud problems. One of them is that violation of Benford's law in many precincts. As statisticians, instead of absorbing the stories from others blindly, we should try to use our expertise and data to figure things out. In class, we have introduced the Benford's law, which is a magic phenomenon in many real-world situations. And in this problem, you are asked to use visualization and analysis under this problem by yourselves.

   In the `CountyData_2000-2020.csv` data set, you will find the county voting counts in the US presidential election from 2000 to 2020. The variable meanings are clear by their names. We focus on Illinois, in which many counties are claimed to have clearly deviation of Benford's law in the 2020 election. Though 2020 and Illinois would be our focus, notice that you can always use the other parts in the data set to help your analysis if necessary.

   1. By visualization, show that whether the first digits of votes in Illinois counties follow/violate Benfold's law. Check this separately for Biden and Trump. (Convincing visualization and comments, 2.5 pts for each candidate)
   2. Do what you observe in Part (a) indicate election fraud? Why or why not? Justify your answer by visualizations. (Effective visualizations 2.5 pts, convincing justification 2.5 pts)

   [Hint: make sure you understand the difference between the two questions above. Crtical logics differences exist. If you feel that Part (a) and (b) are essentially the same questions, perhaps you do not get the point. Our introduction of Benford law in the lecture might be helpful for you to find some clues. Be creative, especially for (b), in designing a way to justify your statements based on all data you could use.]