APPM2720 HW6

---

1) The text data from the *Adventures of Sherlock Holmes* has been cleaned in the class notes as the data set **everything4**. These data should have each work as a separate character string but all the periods are replaced by the string **XXX**.

How many words were found in the document? Make a boxplot of the number of words in each sentence. This book has twelve stories. You can find the beginning of these by searching for **ADVENTURE** in the cleaned data set using the **grep** function.

Use this information to create a figure that has the boxplots of sentence lengths for each story. The idea is that is a simple way to see how different the writing is in each story.

2) The image **CUBoulder.pdf** is an image of the CU Boulder campus printed to file from Google Earth and **CUBoulder.jpg** is an edited version that is smaller and in JPEG format. (Feel free to work with a higher resolution version of the original pdf if your laptop is fast enough.)

Create an algorithm to find the bus stops in this image. Find them by indicating an approximate pixel location for each buss stop.

Create an algorithm to find the roofs of the buildings. Finf these by just coloring all the pixels in the image that look like roofs.

In each case replot the image with your findings.

E.g. If find that a pixel is part of a roof you can just set it to pure white by setting the color channel values all to 1.0. Setting them to the magenta color values is also a good choice to make them stand out

NOTE: Your algorithms may not be completely successful -- that is OK. The most important thing is that your R code is clear, easy to follow and commented.

HINTS: If you are using for loops to analyze the image you might consider adding some print statements inside your loop to help with debugging. Also consider working a much smaller subset of the image first to make things easier to debug.