Stat 419 midterm review session

Material will go through correlation, including Wikipedia notebook. Specifically, the lat and long, and the climate data. The focus of the midterm will be targeted through distance and correlation. Clustering will only focus on some of the most rudimentary forms of clustering. like KMeans and Hierarchical. ( EFA, PCA, SVD will not be included.) Correlation, and distance will be covered. Understanding how to read and interpret correlations from correlations table.

Use Version 1.4.781 to minimize bugginess:

Download from:

https://s3.amazonaws.com/rstudio-ide-build/desktop/windows/RStudio-1.4.781.exe

The Wikipedia notebook you need to make sure you do and that it runs. Go to office hours if needed.

Headings for the midterm:

* Do a basic simulation using Rnorn and Runiform data, and outline what it performs
* Another question tied to easter egg simulation on discussion board.
* A section called "Rolling the dice" , answer questions related to probability.
* Initial exploration of real data: imported from indeed.com, we will be given the data, we'll have to build out some box plots and answer questions about trends
* With computing distances, there will be a question similar to the HW question. We'll have to run the code and change the parameters with a different location.
* Plot state capitals based on latitude and longitude, given various plotting options. Write a response on which one is most beneficial conceptually.
* Can you build a distance matrix of data? (Haversine, Manhattan, etc)
* Next section is hierarchical clustering. What's clustered with what, and does it make sense?
* Must be able to read correlation tables.

The test should take 10 to 15 hours to complete.

Exam will be available at 6am Monday.

All assignments will be graded by 6am Monday.