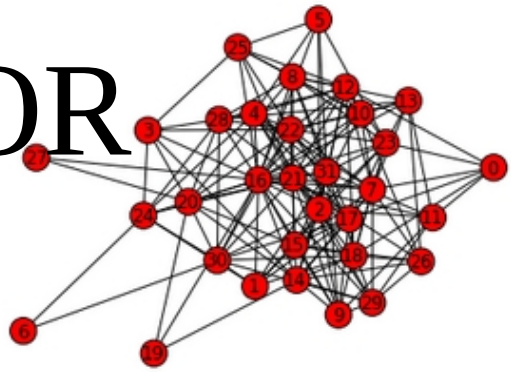
GENERAL ASSEMBLY: Data Science Fall 2014

DATA EXPLORATION07



These are powerful new tools; used today to answer the unanswerable and start life-changing companies. One part math, one part programming, and a whole lot of creativity: data science is an art that cannot be learned from a textbook.

Dataset: Ebola Data. This is a live data aggregated on the spread of the Ebola virus.

1. Subset data and plot deaths and cases by country. Monday's lab

2. Plot cumulative deaths and cases in a stacked bar graph.

3. Use TimeSeriesModeller() to estimate a lagged model for each country what about the TimeSeriesModeller class makes it not ideal for quickly changing epidemic data?

4. Modify the way TimeSeriesModeller() selects the training set so that it fits the model using all data except the last observed datapoint. This involves changing one line of code. What are the tradeoffs with this approach? Refit the models.

5. Create a new data frame with all data transformed by the lograithmic transformation.

6. Use TimeSeriesModeller() to estimate new models on the transformed data. Which is more accurate?