Visual Analytics

Ryan Greenup

May 13, 2020

Contents

Tutorial

- Open Source Packages for Data Viz
- Java Script
  - D3 is a comprehensive JS library utilising SVG HTML5 and CSS to get Data Viz on the web
- Revolution Analytics - R
  - RGL Package is good for 3d in R?
  - HTML Widgets allow building interactive visualisations
- Python

Question 1

Visual analytics is concerned with visual representations of data to facilite reasoning in an analytic fashion. Visual Analytics is usually implemented through a combination of automated analysis techniques relating to data mining, filtering out noise, compressing out unnecessary features and statistical alaysis.

Visual analytics are important in big data analytics because large amounts of data can be unwieldly and difficult to draw insights from without first processing the data. For example a data set with 1, 000, 000 observations over 1000 features will be difficult to interpret despite the fact that there is a sufficient amount of data.

Question 2

Computers can rapidly process very large amounts of information with repeatable results and make high quality graphics (both static and live). Computers cannot however think, it's similar to choreagraphing a chain of dominoes, you can change the layout of the dominoes but you cannot expect the dominoes to alter there own layout.

Humans can solve problems that are difficult to break down into simple steps, simply visualise things that are difficult to compute (e.g. clustering), draw on past knowledge and invent new means to solve a problem.

Ideally the flexibility and inventiveness of a human should be paired with the computer's capacity to efficiently reach repeatable results.

Question 3

Splunk

1. Description Splunk is a visualisation platform used for interpreting big data through a web-style interface, designed to allow access to visualisations across a variety of devices.
2. Pros
   (a) 1. Platform Agnostic *Splunk* implements web technologies to make there software work on a variety of devices, the advantage to this is that you can use a mobile OS to browse visualisations, it can be far more immersive to browse visualisations on an *iPad* than a Desktop for instance.
   (b) 2. Modern Interface The interface to use splunk is a modern web platform which makes for a good user experience.
   (c) 3. Accessible *Splunk* is designed to allow people with out experience in programming to draw insights from visualisations, meaning that anybody can interpret data regardless of there background.
3. Cons
   (a) Slow Interface The problem with using a web interface is that it will be slower and less responsive than a native program, this takes away from the user experience.
   (b) Complex Architecture *Splunk* has a lot of features which means that Implementing it can be quite complex, potentially requiring on-going staff to keep it effective.
   (c) Learning Curve Splunk has a high learning curve, difficult to get started with.

Tableau

1. Description *Tableau* is a commercial visualisation software which can turn data into visualisations through a *GUI*.
2. Pros
   (a) High Performance The unser interface is robust and stable and is renowned to operate fast even on big data.

(b) Large User Base There is a large user base for Tableau which can make finding help easier.

(c) Ease of Use *Tableau* has an easier learning curve than a programming language owing to it's *GUI* interface, this does however have the trade off that there is less room to grow.

3. Cons

(a) High Cost and Training Tableau is commercial software and so it comes attached with a high cost, this means that if it is adopted any visualisations are tethered to the software and that training people to use it will also tether people to that software.
This also means it can be difficult to find people with the training to produce high-quality Tableau Visualisations because individuals may not be willing to undertake training, as of there own volition, for a propriatery piece of software as opposed to something like *GGPlot*, *Plotly* or *D3* which could be used anywhere or any place of emplyment.

(b) Security Issues Tableau is closed-source, this means that the community has had no opportunity to inspect the codebase for vulnerabilities.

(c) Integration Issues It can be difficult to easily embed Tableau into a business's products because of it's proprietary nature, compare this to something like Shiny whcih in contrast can be simpler to implement.

Pentaho

1. Description

2. Pros

(a) Many Features *Pentaho* has a vast array of features that users can implement.

(b) Open Source Components of *Pentaho* are open source, this means that there is a higher capacity for software integration and for community review of software security.

(c) Compatability *Pentaho* is compatable with many database sources including *MySQL* as opposed to for example *Tableau* which requires driver selectors.

3. Cons

(a) Cryptic Error Messages Bugs encountered when using this software are very difficult to fix and the error messages are usually not descriptive of a pathway to rectify the issue.

(b) Error Messages can be Cryptic

(c) Poor Documentation This platform assumes that users are familiar with scripting and the documentation assumes that, so this software can be very challenging to users not experienced with computer programming.

Question 4

Question 5

Question 6