****

**Teaching Notes** for **Humanitarian Uses of Data and Analytics (Powerpoint)**

Provided by Dr. Dave Schrader, [drdaveschrader@gmail.com](mailto:drdaveschrader@gmail.com)

**Motivation:** many college students are interested in “doing good”, applying their skills to help organizations that are typically non-profits and cause-focused. This 50 minute module can be used in analytics classes to generally educate students about the state of the art of donor analytics, and specifically how new data types and systems are being used to help non-profits analyze their data.

There are 3 topics, which should fit into a 50 minute lecture, with some time for discussion.

* The deck begins with a case study about the American Red Cross’s work with Prof. Bill Rand of the University of Maryland to analyze tweet activity for two hurricanes.
* The 2nd topic describes the concept of “Data Dives”, in which non-profits link up with practitioners, who help them analyze their data in ways that perhaps they’ve never thought of before. An organization named DataKind provides the structure for “data for doing good”, and this module comes from a Teradata Partners 2014 DataDive event in Nashville.
* The 3rd topic illustrates the use of Kaggle, a crowd-sourcing approach for companies who have data but perhaps not sufficient analytical talent, to engage data experts from around the world on interesting problems. The examples include new models for epilepsy as well as newer approaches for measuring and improving ocean health.

In addition to the class discussion, you can point students to website information:

* American Red Cross Video on [www.teradata.com](http://www.teradata.com)
* Datakind website: [www.datakind.org](http://www.datakind.org). Also the [www.globalgiving.org](http://www.globalgiving.org) website.
* The Kaggle website which describes many contests: [www.kaggle.com](http://www.kaggle.com)

Slide by Slide background and suggestions for discussion questions:

4 – provides an overview of the some of the newer data types that can now be analyzed, in addition to “traditional” data like donor names, addresses, projects, donation amounts. The deck will provide examples of most of these.

5 – The American Red Cross is a proficient user of data. Traditionally, they have kept information about donors, volunteers (e.g., blood, disaster recovery), and projects in their Teradata system. The nature of the organization is such that they need to do many things in “real-time”. DISCUSS WHY. For example, if there is a tornado and they need type AB-positive blood at a hospital, they need to be able to look up blood donors with the right type who live in the vicinity of the hospital and get the word out. Historically they have done traditional marketing, but with more social media channels, they are exploring new ways of getting the urgent need messages out to the possible volunteers and donors. Media mix analytics as well as media type analytics are important – for example, for donors, is it better to tweet a text message or a picture of a victim? What kinds of pictures get the best reactions (in terms of counts of donors and dollar amounts).

6 - In one research project, ARC worked with the University of Maryland’s Bill Rand (and his students) to explore Twitter activity for two hurricanes, Irene and Sandy. Irene was a minor hurricane; Sandy was devastating. That database included 4M tweets and the main focus was to take a look at the words, the differences in the words used in the tweets, the geospatial distribution of tweets, as well as fundraising results from appeals via Twitter and other channels. DISCUSS how having this information might impact messaging and campaigns.

7&8 – show the counts of tweets over time. You can compare these two to see that Irene tweets were concentrated during the time of the storm, whereas the damage from Sandy was extensive and caused a longer period of tweeting.

9 – in the original presentation by Bill Rand, this was an animation, showing where the tweets occurred over the time duration of each hurricane (use your imagination) DISCUSS how geospatial information could help more tightly focus campaigns, geographically.

10 – shows words used in tweets during one hurricane episode but not the other. Note the tone of tweets for Sandy includes more urgent and graphic words.

11 – shows the timeline for Sandy. The red line at the top shows the duration of the storm. Point out the number of tweets in blue. The amount of donations coming in each day is in green, and the yellow line shows the cumulative total of donations. There are spikes in the donations – they occurred when the American Red Cross fundraising appeals ran on National Public Radio, for example. DISCUSS how this set of information might help refine future campaigns.

12 – the 2nd section of the Powerpoint illustrates the experience that Teradata customers had by participating in an October 2014 “Data Dive” that preceded the Teradata Partners user group event in Nashville. An organization based in NYC called DataKind, led by Jake Porway, organized this event, which was sponsored by Teradata Cares, the philanthropic part of Teradata.

13 – provides some background information about the Data Dives that Datakind has sponsored worldwide. Teradata has been very engaged in the London Data Dives – one of our senior technical people has helped DataKind by soliciting help from the data scientists at companies who use Teradata in the London area, for example.

14 - customers were invited to come 2 days early to the conference to participate in one of 4 Data Dives. DataKind brought in 4 non-profits (and their data). On the first day of the event, representatives from the non-profits gave an overview of their cause and participants decided which group to work with. The rest of the first day was spent asking questions about the data sets, and trying to understand their goals, then formulating “interesting questions” that analysis of the data might help answer.

15 – one of the non-profits is GlobalGiving. If you have a “project” with a fund-raising goal, they will work with you to create a space on their website. DISCUSS “Crowdfunding”. You supply the pictures and text and they will help with the design and the donation buttons (and collecting the money). These pictures shows just a few of the projects – you can have students go to their website to see many more. They have collected hundreds of thousands of donations for 2800 projects and brought that data for our group to analyze.

16 – shows a typical donation page for a project, in this case for Ebola relief efforts. There are interesting analytical questions that we were trying to answer. We split the data set into “successful” and “unsuccessful” projects – each project has a fundraising goal so the criteria was simply whether it had achieved its goals or not. Questions that we wanted to analyze included things like – how many donation buttons should be used? Does it help to have text next to each donation button that says what you get for that amount of money? (In this case, the 3 donation buttons just have the same text – DISCUSS – would it make a difference if you knew what you got for each level of donation?) Do the donation values or intervals matter (e.g., instead of $30, $50, $75 as shown here, would $28, $51, and $78 work better)? Do bigger pictures work better than smaller ones? Do people pictures work better? DISCUSS – what other kinds of analytics questions could be asked?

17&18 – shows our international group of customers and employees who helped analyze the data (along with 2 GlobalGiving executives who explained their data set). On the first day we tried to understand their data models (and the codes for various data elements). We walked through the steps from creating a project to closing it out, including all the steps of donations and how they are coded so we understood the elements of the data that GlobalGiving had collected.

19 – on the second day we worked on variety of the “tantalizing questions” that we’d identified. One question was whether the vocabulary on web pages for successful project descriptions was “different” from the ones that did not achieve their goals. You can see the terms that “worked”. One of the surprising terms was “Goat” – it turns out that for poverty/food campaigns, donations that include buying goats is a winner because they are low-maintenance, can provide milk and meat.

20 – shows our first cut at the analytics for the number of buttons. Start on the right with the number of projects who used various numbers of buttons. The successful ones just had one donation button – freeform where you fill in the blank, but you can see that the average donation amount was lower (so it required more people to donate). DISCUSS: is it better to have more donors, even if they each contribute a smaller amount? How could you decide? (One answer: see if those people donate to more causes than the others – maybe the net $$ amount is the same, or even more).

21 – a 2nd non-profit provides internet-based mentoring opportunities, with a goal of keeping high school kids in school.

22 – this was mostly text analytics, along with timing analytics, to see what kinds of mentor messages “worked”.

23 – summarizes this section by showing that there are DataKind chapters around the world. Students may be interested in finding out more by going to the DataKind website.

24 – the 3rd part of this deck highlights the Kaggle website. This organization helps companies or non-profits who have data but don’t have deep BI analytics capabilities to link up using “gaming” to create contests. Each contest usually has a prize (there are non-prize “knowledge competitions” too), and individuals or teams compete to win the prize (as well as the honor of winning or placing high in each competition).

25 – is an animation of 3 contests just to give a flavor of the kinds of competitions. Many involve “big data” analytics, e.g., looking for face points to analyze human emotion, or analyzing click-through behavior on websites to predict ad success.

26 – show current competitions along with the prize amounts, the number of teams competing, as well as the time duration. The top entry is a non-profit – American Epilepsy Association – that makes its database of brain imagery available with a goal to create new predictive models for epilepsy detection.

27 – as teams create their models, they are scored by the organizers. You can submit multiple models over time to make refinements. You can see here that the winner of the competition improved their model over the time duration of the contest.

28 – shows an interesting new competition, to do image analytics on pictures taken from towed cameras under water, with a goal of automatically counting and identifying types of plankton. This is today a manual and time-intensive operation which Oregon State University’s marine institute would like to automate. DISCUSS: A good homework assignment would be to tell the students to go take a look at and learn about another non-profit contest and come back to tell the class about it next time.

29 – Summary page. Good following quiz questions to ensure understanding:

* What can the American Red Cross do by monitoring tweets during hurricanes?
* What’s a Data Dive? Give two examples of the new kinds of donor analytics are possible?
* What’s Kaggle and why would a non-profit use it?