Alexandra Steinheimer

INFM 600

Info Seeking

Thursday September 15

Data Set 1: Washington Nationals Team Batting Statistics

Baseball Reference.com. (2015). *Team Batting* (Data File). Retrieved from

http://www.baseball-reference.com/teams/WSN/2015.shtml.

The terms of use for this site are located here: <http://www.sports-reference.com/termsofuse.html>. Teachers and students can use the data on the site with proper citation. The terms of use also specify that the data on the site is for informational purposes only.

Two of the characteristics in Davis’ ‘Index of the Interesting’ are ‘Function’ and ‘Evaluation.’ This data is interesting because it can be used to either prove or disprove the team’s success rate and overall effectiveness at playing baseball. If you believe that the only measure of the success of a sports franchise is winning a championship, this data can be used to support or refute that. If you want to argue the level of functionality of 2015 Nationals batters, you can use this data to support your claim as well. Potential users of this data could be baseball fans, coaches, and other baseball players.

This data could answer questions such as:

* Were the 2015 Washington Nationals an effective baseball team? (In this scenario, a definition of ‘effectiveness’ in baseball would be needed.)
* Who were the 2015 Washington Nationals team leaders in batting average?
* What team member on the 2015 Washington Nationals struck out the most?

Data Set 2: NOAA National Weather Service Storm Prediction Center

Storm Prediction Center WCM Page. *U.S. Tornados (1950-2015)* (Data File).

Retrieved from http://www.spc.noaa.gov/wcm/#jmc.

The terms of use for this site are located here: <http://www.weather.gov/disclaimer>. Again, since the data is in the public realm, it can be used for free as long as it is not claimed as one’s own data, it is not modified, and it is not implied that NOAA is affiliated with the person using the data.

According to Davis’ ‘Index of the Interesting’ this data set might be interesting because it might answer questions of ‘Organization’ and ‘Generalization.’ One could analyze this data and make a determination as to whether or not tornados occur at one time or in an outbreak of some sort (or a combination of both). This might be used to prove if a tornado is an unstructured phenomenon or a structured one. This data might also be interesting if it’s used to prove or refute that all Americans face the same tornado threat.

This data could answer questions such as:

* In what region of the United States do most tornados occur?
* During what time of year do most tornadoes occur?
* Are tornados single occurrences or do they mostly occur in an widespread outbreak of violent weather?

Data Set 3: Real Clear Politics Polling Data

General Election: Trump vs. Clinton. *Polling Data.* Retrieved from

http://www.realclearpolitics.com/epolls/2016/president/us/general\_election\_trump\_vs\_clinton-5491.html#polls.

The terms of use of this site are located here: <http://www.realclearpolitics.com/privacy.html>. This data set shows all of the presidential election polls taken in the country; while I assume that the data is free to use with the proper attribution, it might also be advisable to check the terms of use for each poll if data specific to that poll is referenced.

This data set is limited, but is interesting if it can answer questions of ‘Evaluation.’ I would posit that to answer any questions in evaluation of how a campaign preformed in the polls, one would also need an understanding of any outside factors that might have influenced the polling data. Anyone evaluating the data would need to look at news sources from reliable news outlets and attempt to draw a correlation.

Some questions that can be answered with this data are the following:

* What is the average margin of error across the polls?
* What is the average sample size?
* Why might one candidate have a lead in one poll but be trailing in another poll taken during roughly the same time?