

## **Social Data Analysis with StatCrunch: Potential Benefits to Statistical Education**

### **Abstract**

StatCrunch ([www.statcrunch.com](http://www.statcrunch.com)) is an online data analysis package that can be used as a low cost alternative to traditional statistical software for introductory statistics courses. StatCrunch offers a wide array of numerical and graphical routines for analyzing data along with several features such as interactive graphics which can be used for pedagogical purposes. StatCrunch has a number of new features related to social data analysis where users may share data sets and associated analysis results via the StatCrunch site. Users may also interact via online discussions related to shared items. This manuscript provides a brief description of the mechanics of uploading and sharing information via the StatCrunch site and then discusses some of the potential benefits that these social data analysis capabilities offer to both students and instructors.

### **Keywords:**

Social data analysis, StatCrunch

## **Social Data Analysis with StatCrunch: Potential Benefits to Statistical Education**

### **1. Introduction**

The term Web 2.0 has been coined to describe the rapid development of Web sites that encourage users to create and share their own content. This trend in Web development has led to the creation of large Web-based communities around specific topics of interest. For example, Flickr (<http://www.flickr.com>) is a very popular web site dedicated to photos where users can upload their own photos and comment on photos uploaded by other users. YouTube (<http://www.youtube.com>) provides similar functionality centered on video. Other examples of Web 2.0 sites that foster creativity, interactivity and community include the popular social networking sites such as MySpace (<http://www.myspace.com>) and Facebook (<http://www.facebook.com>). Each of these sites has attracted millions of users, and they have fundamentally changed the way that many individuals use the Web.

More recently a number of Web 2.0 sites have also been developed with a focus on the exploration of data. Swivel (<http://www.swivel.com>) is a site that allows users to upload data into a graphics production engine that outputs a variety of graphics for user consumption. Many Eyes (<http://www.many-eyes.com>), developed by IBM, is another site that allows users to upload data, create selected graphics and interact with the graphics. Both of these sites also allow users to discuss their data and associated graphics with other users in a public forum giving rise to what is now called social data analysis. In terms of the basic practice of statistics, the interactive functionality provided by these social data analysis sites is quite exciting because the discussion of data and graphs may in the long run lead to better analyses and better decisions based on data. These sites provide individuals with experience in data visualization with a never ending stream of data to which they can apply their expertise.

Corresponding with the release of StatCrunch 5.0 in August 2007, the StatCrunch web site (<http://www.statcrunch.com>) was also revamped to encourage users to take part in social data analysis. Previously, the StatCrunch site primarily offered students and instructors easy access to online statistical software. The site now has added features that allow for the sharing information in an interactive format. While Swivel and Many Eyes focus entirely on graphics and visualization, StatCrunch provides users with a full suite of numerical and graphical routines for analyzing data covering the entire range of topics typically found in introductory statistics plus much more. This short manuscript will discuss many of the new StatCrunch features and their potential application in the field of statistical education. The discussion will begin with a description of the mechanics related to sharing information on the StatCrunch site and then move on to the benefits that these sharing capabilities offer to both instructors and students. Lastly, some of the broader issues related to the use of social data analysis in statistical education will be discussed, and a few ideas for future enhancements will be provided.

## **2. Sharing information on the StatCrunch site**

Data sets, analysis results and reports can all be shared on the StatCrunch site. This section will serve as a brief tutorial for how this information can be developed and shared. Step by step instructions will be provided covering the entire process from loading and sharing a data set to constructing and sharing reports which combine data sets and analysis results with written discussion. Since the StatCrunch site requires a subscription, a number of large screen shots are used below to help readers understand the basic operations on the site. These screen shots can be accessed by clicking on the links in the figure captions. In many cases the associated information is publicly available in which case the screen shots will also be linked to the live information on the StatCrunch site. Those instructors without a StatCrunch subscription who wish to experiment with the StatCrunch capabilities described below should contact [statcrunch@pearson.com](mailto:statcrunch@pearson.com) about a free instructor subscription.

### **2.1 Sharing data sets**

Both text and Microsoft Excel files can be loaded into StatCrunch from a user's local computer or from a Web address. Figure 1 shows the StatCrunch interface for uploading a local data file. The file being loaded in this case is an Excel file which contains the party and number of votes for each candidate in the 2008 New Hampshire presidential primary. In addition to specifying the location of the file on the local system, there are a number of options available related to sharing the data set. Most importantly, there is a simple share with everyone toggle that can be used to make the data set publicly available. In addition, if the user is a member of one or more StatCrunch groups, options are available for sharing the content with selected groups. Sharing with everyone and sharing with a group need not be mutually exclusive options. The former being a method for placing an item in front of large general audience, and the latter a method for placing an item in front of a more narrow focused audience. Turning off the share with everyone option does not mean an item will not be publicly available. StatCrunch groups can be either public or private. The content of a public group is available to everyone while the content of a private group is only available to the group's members.

There are also entry fields for specifying the original source of the data set as well as a description of the data so that interested parties understand the data characteristics more completely. A tag field is also available which is designed to be a short list of keywords that are relevant to the data set. Tagging is commonly used on Web 2.0 sites as the primary means by which users search information. Users can also specify a thumbnail image to represent a data set. This thumbnail option is designed to provide more visual interest to attract potential users to a data set. The image can be loaded from the user's local computer or by searching the Flickr site within the context of the Web page. Figure 1 shows the results of a Flickr search for images tagged with the term "new hampshire primary". A specific image from the Flickr search results can be selected by simply clicking on the image.

## Load data from my computer

**File:**

**Use first line as column names:**

**Delimiter:**  
   
(required for text files only)

**Share with everyone:**  No  Yes

**Share with groups:**

|                          |  |
|--------------------------|--|
| <input type="checkbox"/> | StatCrunch help  |
| <input type="checkbox"/> | Using StatCrunch to broaden the horizons of an introductory statistics |

[Learn more about sharing with groups!](#)

**Source: (optional)**


**Tags: (optional)**

### Data Set Properties

[[Edit](#) - [Delete](#)]


---

**Thumbnail:**



from Flickr

**Owner:** [websterwest](#)



**Last edited:** Jan 16, 2009

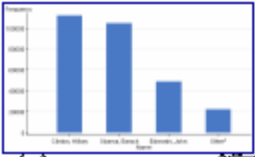
**Size:** 1KB

**Share:** yes

**Views:** 1108

**My results for this data:**

NH Democratic Primary Results



### NH Primary Data

[Link](#) [Embed](#)

StatCrunch
Data
Stat
Graphics
Help

| Row | Name             | Party | Votes  |
|-----|------------------|-------|--------|
| 1   | Clinton, Hillary | Dem   | 112238 |
| 2   | Obama, Bara>     | Dem   | 104757 |
| 3   | Edwards, Jo>     | Dem   | 48666  |
| 4   | Richardson, >    | Dem   | 13245  |
| 5   | Kucinich, De>    | Dem   | 3912   |
| 6   | Total Write-ins  | Dem   | 3076   |
| 7   | Biden, Joe       | Dem   | 628    |
| 8   | Gravel, Mike     | Dem   | 402    |
| 9   | Caligiuri, Ric>  | Dem   | 254    |
| 10  | Dodd, Chris      | Dem   | 202    |
| 11  | Capalbo, Ke>     | Dem   | 106    |
| 12  | Hunter, D.R.     | Dem   | 95     |
| 13  | Keefe, Bill      | Dem   | 47     |
| 14  | Laughlin, Tom    | Dem   | 46     |
| 15  | Crow, Randy      | Dem   | 35     |

**Description:**  
This data set contains the number of votes received by e

**Source:**  
CSPAN, <http://www.cspan.org/>

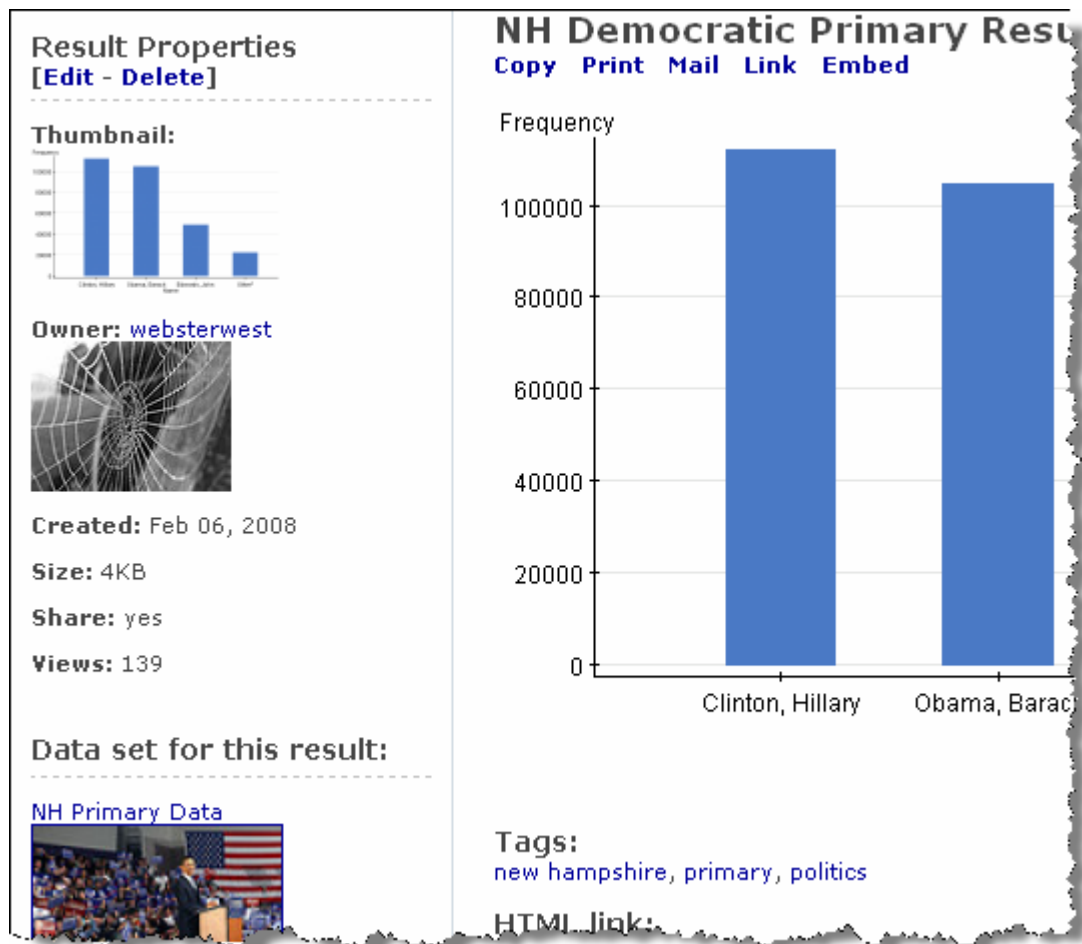
**Tags:**  
[new hampshire](#), [primary](#), [politics](#)

Figure 2: See <http://www.statcrunch.com/jse/figure2.jpg> for the complete image

## 2.2 Sharing results

StatCrunch has a number of numerical and graphical procedures available for analyzing data. While this manuscript focuses on the sharing capabilities of the StatCrunch site, West et. Al. (2004) provides more information on analyzing data with StatCrunch. Analysis results in StatCrunch can be saved to the user's My Results listing by using the *Export to My Results* option under the *Options* menu of the result frame. StatCrunch results are exported in HTML format so that graphics and tables are well formatted for easy copying, pasting and printing. Figure 3 shows a bar plot of the results from the New Hampshire democratic primary as exported to the My Results folder. In this case, the bars have been ordered in decreasing fashion in terms of the number of votes received by the candidate, and those candidates with less than 5% of the vote have been grouped into

an *Other* category. Complete details on constructing this result with StatCrunch statistical software are provided in the report described below.



**Figure 3:** See <http://www.statcrunch.com/jse/figure3.jpg> for the complete image

When a result is exported to My Results, it is by default not shared. This result property can be changed by clicking the Edit link on the result page. The full set of editable result properties are shown in Figure 4. As with data sets, there is a simple no/yes toggle for sharing the result and searchable tags can be added to the result. When a result is exported, it has a simple descriptive title based on the type of analysis being conducted. For example, the “Histogram” title is assigned to all histograms when they are exported. Before sharing a result, one can change the title of the result to be more descriptive in order to capture the interest of other users. The user can also specify notes for a result which is designed to be a short textual description of the result or insights gained from it. If the result contains a graphic, a resized version of the graphic will automatically be used as the thumbnail image to represent the result on the StatCrunch site. If the result does not contain graphics or the user wishes to change the default graphic, then an image file can be uploaded or a representative image can be selected from a search of the Flickr site.

## Result information

**Title:**

**Share with everyone:**  No  Yes

**Share with groups:**

StatCrunch help

Using StatCrunch to broaden the horizons of an introductory statistics

Learn more about sharing with groups!

**Tags: (optional)**  
  
 If you share your result, these terms will make it easier to find via search. Separate tags with spaces.

**Notes: (optional)**

**Figure 4:** See <http://www.statcrunch.com/jse/figure4.jpg> for the complete image

### 2.3 Sharing reports

StatCrunch now offers the ability to write complete online reports where a discussion of data analyses can be supplemented with embedded data sets and associated analysis results. New reports can be created from a user's My StatCrunch page which also provides a listing of the user's current reports. Figure 5 shows the options available when creating a report. As was the case with data sets and results, reports can easily be shared using a basic no/yes toggle and reports can be tagged with relevant keywords. Users can also specify an image to represent the report on the StatCrunch site by loading an image from their local computer or by searching Flickr. Basic HTML syntax or plain text can be used to construct the body of the report. Data sets and results to be embedded within a report can be selected using simple list selectors with associated numbers. The user can select results and datasets one through five by default when creating a report. If more results or data sets are required, the user can simply save the report and then reopen it in which case five additional result or data set numbers will be added to the report. By default, results and data sets selected by the user will be tacked on at the end of the report. They can also be manually positioned by entering tags like `<result#>` or `<data#>` within the body where # is replaced with associated result or data set number. For example, to include the first attached result at a specific location in the body of the report, the tag `<result1>` is entered at the desired location. If the user chooses to share a report, the data sets and results that are attached to the report do not have to be shared in order to be visible by other users in the context of the report. In other words, if these attached items are not shared, they will be visible when someone views the shared report but will

not be available to other users as individually shared data sets or individually shared results.

**Report information**

**Title:**  
Creating bar plots from summary data

**Share with everyone:**  No  Yes

**Share with groups:**

StatCrunch help

Using StatCrunch to broaden the horizons of an introductory statistics course

[Learn more about sharing with groups!](#)

**Tags: (optional)**  
"bar plot" "summary data"  
If you share your result, these terms will make it easier to find via search. Separate tags with spaces.

**Body:**  
Editor: [Add/Remove](#)

**B I U** | ABC [List icons] [Undo] [Redo] [Link] [Image] [Table] [Equation] [HTML] [Help]

Summary data in StatCrunch is used to refer to data sets that do not consist of... but rather data summarized from other sources. The attached New Hampshire... of summary data. In this case, we do not have the results for all of the individu... rather we have the summary totals for each candidate in the democratic and re...

<data1>

This report will discuss how to make bar plots with summary data using the Ne... as an example. The steps needed to construct a bar plot for the results from th... discussed. Begin by selecting the Graphics > Bar Plot > with summary option. M... column for the Categories in selector and the Votes column for the Counts in sel... the democratic party results, input Party = GOP in the Where text box as shown...

<result1>

Click the Next button to specify more options. To combine the numbers for cand... percentages (say less than 5%) so that they do not obscure the graph, specify

**Figure 5:** See <http://www.statcrunch.com/jse/figure5.jpg> for the complete image

The report being created in Figure 5 contains the instructions for creating the bar plot shown in Figure 3 with StatCrunch software. The resulting report is displayed in Figure 6. Using the Snapshot feature in the StatCrunch statistical software, a user can take a

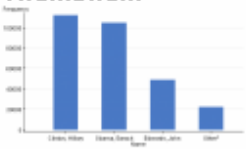
screen shot of the dialog windows used to perform a specific analysis. These snapshots are included in the user's My Results folder and can then be included within reports. Figure 6 shows a number of these snapshots as part of the instructions to create the bar plot shown in Figure 3. The snapshot feature makes it easy to develop reports containing detailed instructions for specific analysis techniques. As shown in Figure 6, data sets are embedded in reports in analyzable format within StatCrunch software. Readers can then analyze the data being discussed within report to arrive at their own conclusions based on their own assumptions.

### Report Properties


[Edit - Delete]

---

**Thumbnail:**



**Owner:** websterwest



**Created:** Feb 25, 2009

**Share:** yes

**Views:** 613

**Tags:** bar plot, summary data

---

**Results in this report**

1. Snapshot of Bar plot with summary Dialog
2. Snapshot of Bar plot with

## Creating bar plots from summary

[Mail](#) [Print](#)

Summary data in StatCrunch is used to refer to data but rather data summarized from other sources. The of summary data. In this case, we do not have the re rather we have the summary totals for each candidat

**Data set 1. NH Primary Data** [\[Info\]](#)

[StatCrunch](#) [Data](#) [Stat](#) [Graphics](#) [Help](#)

| Row | Name             | Party | Votes  |
|-----|------------------|-------|--------|
| 1   | Clinton, Hillary | Dem   | 112238 |
| 2   | Obama, Bara>     | Dem   | 104757 |
| 3   | Edwards, Jo>     | Dem   | 48666  |
| 4   | Richardson, >    | Dem   | 13245  |
| 5   | Kucinich, De>    | Dem   | 3912   |
| 6   | Total Write-ins  | Dem   | 3076   |
| 7   | Biden, Joe       | Dem   | 628    |
| 8   | Gravel, Mike     | Dem   | 402    |
| 9   | Caligiuri, Ric>  | Dem   | 254    |
| 10  | Dodd, Chris      | Dem   | 202    |

**Figure 6:** See <http://www.statcrunch.com/jse/figure6.jpg> for the complete image

## 2.4 A user's public profile

The complete collection of data sets, results and reports being shared by a user can be accessed under the user's public profile. The public profile is available as a link from the user's My StatCrunch page. The basic form of the link is **<http://www.statcrunch.com/profile.php?id=userid>** where *userid* is replaced with the user's StatCrunch id properly urlencoded. This link can be embedded within documents to point readers directly to a user's shared information. For example, to access the author's public profile simply [click here](#).

**Public profile for websterwest**

Shared data sets | Shared results | Shared reports

Showing 1 to 107 of 107 data sets

**Data Set/Description**



**Responses to Big cities in America survey**  
 This data set consists of responses from a survey offered on the statcrunch.com site from 3-6-2009 to 3-13-2009. Responders were asked to select cities from a list of the 15 largest US cities that they would like to visit and most/least like to live.



**US music album sales by genre from 2006 to 2008**  
 This data set contains the yearly album sales figures by genre from 2006 to 2008.



**Responses to How often do you text?**  
 This data set was collected via a survey on the statcrunch.com site from 2-26-2009 and 3-6-2009. Responders were asked to provide the number of text messages that they send and receive each day along with their responses.



**Traffic congestion and housing prices in 2008**

**Figure 7:** See <http://www.statcrunch.com/jse/figure7.jpg> for the complete image

An example of a public profile is shown in Figure 7. A user's public profile contains limited information about the user including their occupation (faculty, student or professional), their associated organization and an identifying thumbnail if the user has specified one. The profile also lists the most recently shared data sets, results and reports with links to the complete listings. A StatCrunch subscription is not required to view these shared items. A subscription, however, is required in order to analyze data using StatCrunch software.

### 3. Using StatCrunch sharing capabilities to support statistics education

The sharing capabilities outlined in the previous section can be used for many statistical education objectives. The most obvious pedagogical application is that instructors can use the StatCrunch site to share a large number of data sets with students in a very convenient fashion. This capability can help instructors more easily use large real data sets as called for in the Guidelines for Assessment and Instruction in Statistics Education ([GAISE](#)) published by the American Statistical Association. By linking to their public profile, an instructor can provide students with one click access to data and the software to analyze it.

**View options**

**Browse all**

politics Search

**Popular data tags:**

AGE  
ANALYSIS  
ASSESSMENT  
BASEBALL  
BOOK  
CAR  
CEREAL  
CHAPTER  
CLASS  
CLASSROOM  
COLUMN

**Data sets shared by StatCrunch**  
Showing 1 to 6 of 6 data sets matching **politics**

**Data Set/Description**

**Global politics 23.xls**  
[Image of a globe]

**Responses to the StatCrunch elec...**  
This data set was collected via a poll...  
poll was voluntary response, please u...

**NH Primary Data**  
This data set contains the n...  
candidate in the New Hamp...  
primaries.

**Figure 8:** See <http://www.statcrunch.com/jse/figure8.jpg> for the complete image

Instructors can also use the StatCrunch site to search for interesting data sets to use in their classes. A complete searchable listing of all data sets being shared on the StatCrunch site can be obtained by clicking the *Data* link at the top of each StatCrunch page. As an example, Figure 8 shows the results of a search of shared data sets for the term “politics.” A list of commonly used tags for data sets is also available which can be used to quickly conduct searches of interest. Once a data set of potential interest has been identified, an instructor can evaluate the data set for characteristics of interest by referring to the list of results and reports for the data set being shared by StatCrunch users. See the left hand side of Figure 3 for an example. This feature can save an instructor a great deal of time as they can peruse these shared results instead of regenerating the results. Surveying the results shared by other users may also quickly lead the instructor to more interesting analyses which can be made with StatCrunch software. Students may also benefit from the data set sharing capabilities offered by the StatCrunch site as they can locate data examples that are of interest to them rather than relying on the instructor as their sole source of data. Indeed both students and instructors should feel awash in data as they use the StatCrunch site.

A searchable list of all shared results on the StatCrunch site can be obtained by clicking the *Results* link at the top of all StatCrunch pages. Both students and instructors may find searching this listing of interest. Students may wish to apply knowledge gained from their course by searching for different types of results. For example, a student may

search on the term “histogram” as shown in Figure 9 in order to gain more practice interpreting shape. In this case, the search yielded more than 1600 histograms which the student can than page though to find examples of different shapes. Instructors may find this facility useful as well as they search for different items to include on exams and other assignments. Most instructors have had the desperate feeling of looking for a right skewed histogram to include on an exam. With the StatCrunch site, one can be found quickly even without a subscription. A link to the associated data set (if it is being shared) is available on each shared result page so that user’s can trace back to the original data and other shared results. Thus, there are a number of avenues by which students and instructors can find their way to items of interest to them.

The screenshot shows the StatCrunch search results for the term "Histogram". On the left, there is a sidebar with "View options" and a list of statistical methods including ANOVA, Bar plot, Boxplot, Chart Group Stats, Chi-Square, Column Statistics, Contingency Table, Correlation, Covariance, Dotplot, Frequency Table, Histogram, Index Plot, Kruskal-Wallis, and Logistic Regression. The main area displays "Results shared by StatCrunch" showing 1 to 15 of 2393 results. Four histograms are visible in the results list:

- Histogram**: A histogram with a peak around 1500.
- Histogram**: A histogram with a peak around 20.
- Alexa's Histogram**: A histogram with a peak around 15.
- Histogram of Student Ages**: A histogram with a peak around 20.

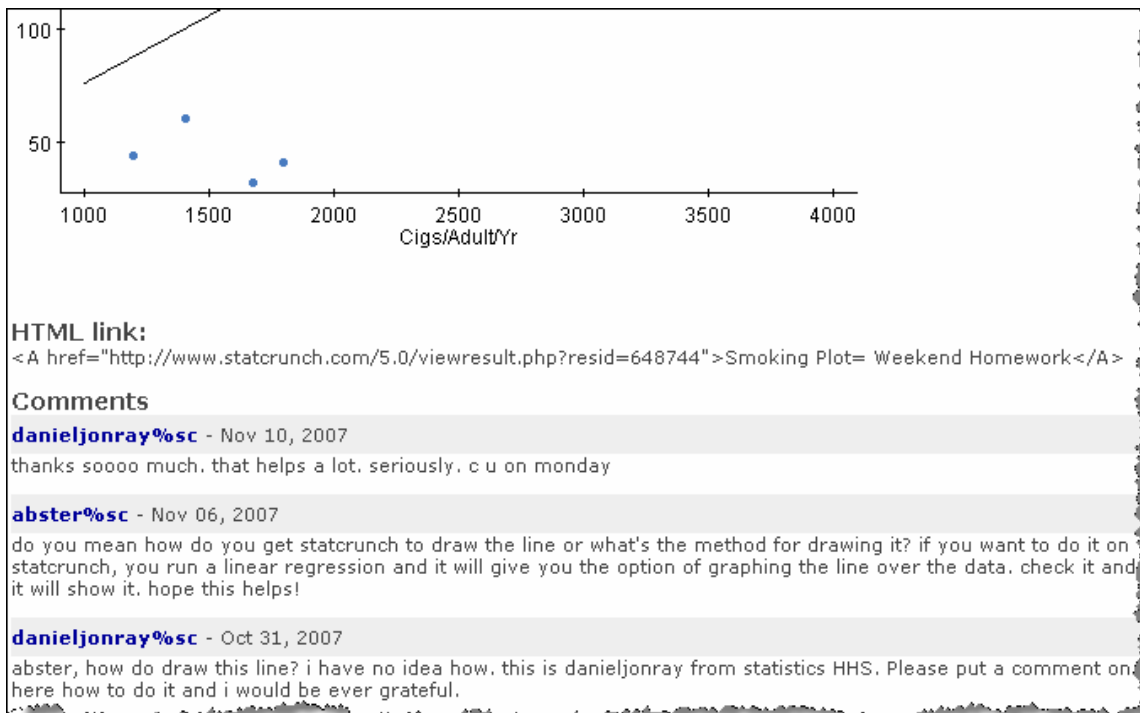
**Figure 9:** See <http://www.statcrunch.com/jse/figure9.jpg> for the complete image

The sharing of StatCrunch reports also has some clear educational applications in courses where students turn in lab reports or data analysis projects. Each of these items can be developed as a StatCrunch report which can then be shared with an instructor as well as other students. Students can email the instructor a notification about a report with a link to it. In addition to being a convenience to the instructor, students can also learn a great

deal from observing and evaluating the work of other students. In order to make reports easier to find by students in the same course, students should be encouraged to add tags to their reports that in some way represent their course. For example, students might tag their reports with the last name of their instructor, the StatCrunch id of their instructor or their school name. Some instructors might naturally be concerned that students would share results when creating StatCrunch reports for project purposes, and thus that it would be impossible to determine who did the work. However, each result embedded in a report includes the StatCrunch id and the time the result was created so that an instructor can at least be confident that the same StatCrunch account used to create a report was also used to create the results included in it. This is in fact much more protection from cheating than one would have from a traditional project report done in a word processor with components that can be readily copied and pasted between students.

The StatCrunch site offers the “social” aspect of the term social data analysis by providing tools that allow users to communicate with one another. Each web page for a data set, a result or a report has a comment listing at the bottom. StatCrunch subscribers have the option of adding comments to each of these pages. The ability to add comments is restricted to subscribers in order to police malicious comments made by spammers. Each time a comment is posted on a particular page, an email is sent to the owner of the item so that they may respond in kind. Perhaps the biggest benefit of conducting these online conversations is the fact that the comments are documented for the rest of the world to see. Most instructors realize the futility of email as a means of answering technical questions as they often answer the same question numerous times as posed by a number of students. These online public conversations, however, allow others to directly benefit from the interaction.

The commenting tool can be used to facilitate instructor-to-instructor, student-to-instructor and student-to-student interactions which can benefit the pedagogical process. Instructors who teach statistics without a great deal of formal training can use the commenting tools establish connections with more seasoned instructors. They may pose questions to other instructors about the specifics of how to use a data set in a course, the appropriate methods to analyze a particular data set or the interpretation of a result. This type of interaction may over time lead to the establishment of instructor networks that lead to improvements in statistical education across institutions. Instructors may also choose to interact with students on the StatCrunch site. For example, if a student posts a shared report on the StatCrunch site, an instructor might choose to provide some written feedback (no including a grade of course) to the student using the commenting tool. The benefit being that other students can learn from these comments when viewing the shared report. Students interacting together have made up most of the interactions to date on the StatCrunch site. In many cases, these online student collaborations have been used to solve problems. For example, Figure 10 highlights a very short interaction between two students where one student helps the other student add a regression line to a scatter plot in StatCrunch. An interesting application of student-to-student interaction would be to ask students to peer review the StatCrunch work of others before turning in final versions of their work. This type of approach has been commonly used in the traditional class room and translates well into the online environment that the StatCrunch site provides.



**Figure 10:** See <http://www.statcrunch.com/jse/figure10.jpg> for the complete image

#### 4. Discussion

The idea of sharing data sets for teaching purposes dates back to the DASL (<http://lib.stat.cmu.edu/DASL/>) project. The recent additions to the StatCrunch site advance on these ideas by allowing users to easily add data sets and to interact around a variety of shared items. Traditional statistical packages such as Fathom, JMP or Minitab typically are distributed with a short list of standard data sets whereas the list of data sets at the StatCrunch site continues to grow at an impressive rate. Even other online utilities such as SOCR (<http://www.socr.ucla.edu/>) also come with a small fixed number of data sets. The sharing capabilities at the StatCrunch site are also quite different than those of other online sites like Wiggio (<http://www.wiggio.com>) or Google Apps Education Edition. These sites provide general tools for sharing information like calendars and documents. The StatCrunch site, on the other hand, is different in that it provides the capability to create items on the site. Indeed, using the StatCrunch statistical software to author items that are subsequently shared on the site is one of its most unique features.

While the social data analysis capabilities described above have only existed on the StatCrunch site for the few short months, numerous items have been shared on the StatCrunch site with most of the shared items coming from students who are completing assignments. As one might expect, many of these items go without comment. In fact, only a very small proportion of shared items have received comments to date. This may in part be due to the fact that the educational users of the StatCrunch site are unaware of some of the potential benefits that online discussions provide. This possibility served as one of the primary motivations behind this manuscript. One of the main goals for the

StatCrunch site in the coming months will be to establish much more of an interactive community amongst the instructors and students that use the site.

One of the key components to developing a large community of users for many successful Web 2.0 sites has been to heavily reward early adopters. Along these lines, regular (monthly or quarterly) awards for the most interesting new data set, result and report are being considered for mid 2008 with a panel of StatCrunch instructors charged with making these determinations. The panel will also be charged with providing feedback in the form of comments about the top candidates in each category. Also being considered is the posting of monthly challenge data sets on timely topics that will hopefully entice the interest of many users.

Advancements in technology may also play a large role in the development of user communities at the StatCrunch site. The goal for student users is to make their usage driven more by desire rather than by duty in terms of completing assignments. Also geared for promoting user interaction, user developed groups designed around very specific topics of interest have recently been added to the site. The creator of the group serves as moderator and controls the flow of new members into the group. Members of the group can post their shared items (data sets, results or reports) on the group's web page for discussion. User groups may work quite well for statistics courses with the instructor being the moderator and students being the members. This will allow students to easily identify content developed and shared by their instructor or fellow students. Other additional features such as item ratings are also being developed so that users can more easily identify quality items of interest that will hopefully inspire online interaction.

If the above strategies for cultivating online interactions are successful, monitoring the quality of information posted by users will always be an ongoing challenge. One malicious user can post a large number of misleading items or comments that impact a large number of other users. One method for dealing with this issue that is currently being considered is the development of an instructor certification program. This program will not in any way seek to measure the knowledge of an individual instructor but rather it will certify that the user is an instructor for a statistics course. The steps for certification will not be overly rigorous requiring only a short online application with user specified links verifying that the user is in fact a statistics instructor. Another strategy for dealing with the issue that is also being considered is a comment rating system where individuals rate the usefulness of comments made by other users. This will allow users to judge the quality of a particular comment based on historical evidence. Indeed, links to highly rated users and their comments may in fact be featured on the StatCrunch site as sources of quality information.

The ideas expressed above for using social data analysis to support statistical education are by no means exhaustive. Technological advancements will also undoubtedly bring new possibilities in the future. The StatCrunch developers are always interested in ideas from users in this regard. Feedback based on user experiences will be a vital integral part of the future development of both the statistical software and the StatCrunch site capabilities. User feedback will also hopefully allow for more detailed research on how these capabilities impact student learning.

## References

Google Apps Education Edition. Available online at:

[http://www.google.com/educators/p\\_apps.html](http://www.google.com/educators/p_apps.html).

Guidelines for Assessment and Instruction in Statistics Education (**GAISE**) for the College Introductory Course. Available online at:

<http://www.amstat.org/education/gaise/GAISECollege.htm>.