PRESS BOOK

**Part I.DATA CLEANING AND CONVERSION**

**Choosing the Right Dataset**

(April11, 2015, Linghong)

The security and safety data has on campus, residence halls, non-campus and public property four groups of dataset. However, the security data for non-campus and public property have low number, which makes difficult to distinguish whether a change is a random variation or a real pattern. For example, weapon possess is one of the data set that has low crime number. From the four total weapon possess charts (fig1 a-d), we can see three of these charts have somewhat downward trend, but fig 1c has very low crime number and a high random variation, which makes difficult to discover any pattern or trend.

Fig1a total illegal weapon possess happened on campus.

Fig1b total illegal weapon possess found in college residence halls

Fig1c total illegal weapon possess found in non campus area

Fig1d total illegal weapon possess found in public property

The above charts also provides a clear indication that compared other groups of dataset, on campus data has relatively high number. Research further of the other data groups that has low numbers gave us similar conclusion. Also considering that in some school categories (such as community colleges), students usually don't live in campus resident halls, thus comparing the data in residence halls could cause bias, hence, we decide to only focus on the on campus security and safety data.

The on campus data has arrest, crime, discipline and hate four degree of violations. Arrest and discipline has three crime categories, while the other two have nine categories with arrest and crime categories have more severe situation than other two categories, thus for a project with limited time frame, we decide to make a site only focusing on campus arrest and crime.

**Data Cleaning and Conversion**

(April 11, 2015-April 16,2015, Linghong and Suhas)

We first combined on campus arrest and on campus crime data into one file. They were totally 12 categories. Each file covers three year period. Then all schools whose state columns are empty were deleted. Those schools are the oversea schools established by US colleges. Further analyzing the data, we found the data also includes colleges in Puerto Rico and US virgin islands. They have state name, hence we didn't notice in the beginning. Those schools thus were deleted from our data too. About 10% of colleges didn't submit security data. Those schools were also deleted from our list. The clean data were then linked to geo coding number (see below) and converted to tsv files.

**Connect Campus Security Data to Geo Data**

(Date: Peter)

**Part II. DATA ANALYSIS**

**Data Analysis Find Interesting Patterns and Trends**

(April 4.14-4.17, Linghong)

The initial try of analysis of the data didn't turn into precise result because we didn't notice there were schools belong to Puerto Rico and US Virgin Islands. We use the clean data reanalyze the data again. Since each category has very different number of schools and students, we use total school number or total students number in each school category to normalize the data. The data analysis provides a clear indications of following trends:

**I. On Campus Arrest and Crime Number Is School Sector Related**

The edu.gov site provides 9 years of security data. To save time, we only did an analysis of the most recent 6 years for 6 school sectors that are generally more interested by students, parents and teachers.

Research of the on campus arrest data showed a clear crime pattern related to school sectors(Fig2a-c). Generally speaking, public 4 year colleges have much higher number of on campus crime issue compared to other 5 college categories. Since the school number between each school category varies significantly, the data shown here are normalized by school number, and listed as crime case number per 100 schools in each school sector. The school sectors are: 1: Public, 4-year or above, 2: Private nonprofit, 4-year or above, 3: Private for-profit, 4-year or above,4: Public, 2-year,5: Private nonprofit, 2-year,6: Private for-profit, 2-year.

fig2a. On campus arrest for weapon violation during 2008-2013

fig2b. On campus arrest for drug violation during 2008-2013.

fig2c. On campus arrest for liquor violation during 2008-2013.

To analyze the nine categories of on campus crime data, we selected six categories that have relative large number for analysis. Interestingly, we also see a distinctive pattern that public 4 year schools have much higher crime number compared to other five college sectors(fig2a-f). Same as above, the listed number are shown by cases number per 100 schools in each school sector. The school sectors are as below: 1: Public, 4-year or above, 2: Private nonprofit, 4-year or above, 3: Private for-profit, 4-year or above,4: Public, 2-year,5: Private nonprofit, 2-year,6: Private for-profit, 2-year

fig3a. On campus crime for forcible sex offense between 2008-2013.

Fig3b On campus crime for robbery violation during 2008-2013.

Fig3c On campus crime for Aggravated Assault during 2008-2013.

Fig3d On campus crime for burglary during 2008-2013.

Fig3e On campus crime for vehicle theft during 2008-2013.

Fig3f On campus crime for arson violation, 2008-2013.

This above school sector related crime pattern is across all the nine on campus crime and arrest categories we have done the analysis, thus we believe our finding is real. We also assumed that the likely same pattern would exist in the state level. **We thus decide one of our box will be used to display the relationship between school sector and crime safety data.**

**II. Yearly Trend: Public School Sector Gets Better, Other School Sectors Have No Change**

Further analysis also displayed an interesting trend of public school crime data: among nine categories that we did analysis, six categories have a downward trend of crime number in the recent years. In contrast, there is no visible change seen in other school sectors.(Fig4a-i)

Fig4a On campus arrest for weapon violation during 2008-2013. There is a trend of decrease of crime number in public 4-year colleges.

Fig 4b On campus drug violation during 2008-2013. Unlike the other category, public 4-year colleges have no improvement in drug violation.

Fig4c On campus liquor violation during 2008-2013. Public four year colleges have a trend of decrease of violation number.

Fig4d On campus forcible sex offense data showed that public four year colleges have a trend of increase of violation number.

Fig4e On campus robbery violation during 2008-2013. Public four year colleges display a trend of decrease of violation number.

Fig4f On campus agg violation during 2008-2013. Public four year colleges have a trend of decrease of violation number.

Fig4g On campus burglary during 2008-2013. Public four year colleges have a trend of decrease of crime number.

Fig4h On campus vehicle theft during 2008-2013. Public four year colleges have a trend of decrease of theft number.

Fig4i On campus arson violation during 2008-2013. Public four year colleges have a trend of decrease of violation number.

Despite we didn't see all on campus security categories have downward trend in public 4 year college sector, we still believe it is an interesting finding , and a nice compensation to the finding that public 4 year colleges have high per campus crime data. **Thus , if time is granted,** **we may implement the yearly crime trend and display in one of our boxes.**

**III. Possible Student Number Related School Sector Pattern**

We also did analysis by normalizing the security data with total students in that college category. The results also displayed a distinctive pattern. In short, the three campus arrest groups have a pattern like fig 5a, while various campus crime data have a pattern like fig 5b. More importantly, for all of these patterns, we see a consistent high crime number in category 5 across various crime categories. (1: Public, 4-year or above, 2: Private nonprofit, 4-year or above, 3: Private for-profit, 4-year or above,4: Public, 2-year,5: Private nonprofit, 2-year,6: Private for-profit, 2-year)

fig5a. on campus drug violation happened during year 2008-2013. Data are listed by crime cases per 10000 students

Fig5b on campus burglary during 2008-2013. Data are shown crime cases per 10000 students

One thing we have to mention is that the student number, female and male number listed in the dataset is not for each university campus, instead it is for each university, however, the security data provided is for each campus. Since not all university campus provides security data, the way of normalization by total student number could produce imprecise result. College sectors with more universities campus that didn't submit security data could benefit of this normalization and display artificially lower data. Thus, we think the campus security data do show some possibility of school size related school sector pattern, but have to be implicated with cautious.

Because of that, we won't implement this finding into our site. For the same reason , we also won't research further for the possible relation between the school size/female male ratio and the campus security in our site.

**PART III WEBSITE**

**Site Structure and Responsive Design**

(April 8-9, April 11 Linghong)

The site main structure is a map that function as visually displaying college on campus security data. Alongside the map, there are a few boxes, which act as either a user interactive area to control the map view, or a place to show the college security patterns or trends. The boxes size, color or position will be changed according to the content size and visualization purpose.

The whole site (including the map) is made responsive to screen size change, and can be viewed in desktop, i-pad and smart phone with a layout suitable to each screen size. Especially, when viewing the site in a smart phone, the menu button will be collapsed. Clicking the button, the menu will show up. Bootstrap are used to organize the site structure, look and the responsive design.

**Map View**

(Date, Suhas)

**Interactive Control Panel and Weighting Method**

(Date, Suhas)

4/22/2015

All visualizations will get the data from data/crimedata.json. This file will have raw crime data in the following structure:

A small json sample is available in data/sample.json.

{**"type"**:**"crimeData"**,**"schools"**:[  
{  
 **"school"**: {  
 **"schoolId"**:**"100654001"**,  
 **"name"**:**"Alabama A & M University"**,  
 **"branch"**:**"Main Campus"**,  
 **"address"**:**"4107 Meridian St"**,  
 **"city"**:**"Normal"**,  
 **"state"**:**"AL"**,  
 **"zip"**:**"35762"**,  
 **"latitude"**:**"34.733964"**,  
 **"longitude"**:**"-86.522913"**,  
 **"sectorCd"**:**"1"**,  
 **"yearData"**:[  
 {**"yearOfData"**:**"2008"**,  
 **"murderCount"**:**"0"**,  
 **"negligentManSlaughter"**:**"0"**,  
 **"forcibleSexOffense"**:**"1"**,  
 **"nonForcibleSexOffense"**:**"0"**,  
 **"robbery"**:**"4"**,  
 **"aggravatedAssault"**:**"26"**,  
 **"burglary"**:**"122"**,  
 **"vehicleTheft"**:**"3"**,  
 **"arson"**:**"1"**,  
 **"weaponOffence"**:**"4"**,  
 **"drugViolations"**:**"9"**,  
 **"liquorViolations"**:**"1"**},  
 {**"yearOfData"**:**"2009"***/\* rest of the fields will repeat here\*/*},  
 {**"yearOfData"**:**"2010"***/\* rest of the fields will repeat here\*/*},  
 {**"yearOfData"**:**"2011"***/\* rest of the fields will repeat here\*/*},  
 {**"yearOfData"**:**"2012"***/\* rest of the fields will repeat here\*/*},  
 {**"yearOfData"**:**"2013"***/\* rest of the fields will repeat here\*/*}  
 ]  
 }  
 }  
 ,{**"school"**:   
 {**"schoolId"**:**"100663001"***/\*school data will repeat here\*/*}]}}  
]}

To process this raw data, a new js file crimedataAnalyzer.js is created. The constructor of this file received the raw json data and it is expected to process all “views” of the data for the individual parts of the visualizations.

For example, the map requires the data aggregated by year and data for all years. In the init method, it populates yearCrimeData and allTimeCrimeData.